

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2006, 10:13:30 ; Search time 297.073 Seconds
(without alignments)
29.581 Million cell updates/sec

Title: US-09-606-910E-2

Perfect score: 109

Sequence: 1 HGVTSPAPTRPAGSTAPPA 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_21.*

1: Geneseqp1980s.*

2: Geneseqp1990s.*

3: Geneseqp2000s.*

4: Geneseqp2001s.*

5: Geneseqp2002s.*

6: Geneseqp2003as.*

7: Geneseqp2003bs.*

8: Geneseqp2004s.*

9: Geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	109	100.0	20	2	AAY25112 Synthetic
2	109	100.0	20	5	ABB08278 Human can
3	109	100.0	20	9	ADX58123 Cancer-as
4	109	100.0	20	9	AEB96326 MUC1 O-li
5	109	100.0	21	2	AAY25111 Synthetic
6	109	100.0	21	3	AAY96171 MUC1 repe
7	109	100.0	21	3	AAY96173 MUC1 repe
8	109	100.0	21	3	AAY96169 MUC1 repe
9	109	100.0	21	3	AAY96180 MUC1 repe
10	109	100.0	21	3	AAY96179 MUC1 repe
11	109	100.0	21	3	AAY96172 MUC1 repe
12	109	100.0	21	3	AAY96174 MUC1 repe
13	109	100.0	21	3	AAY96176 MUC1 repe
14	109	100.0	21	3	AAY96178 MUC1 repe
15	109	100.0	21	3	AAY96177 MUC1 repe
16	109	100.0	21	3	AAY96175 MUC1 repe
17	109	100.0	21	8	ADK65929 MUC1 glyc
18	109	100.0	21	8	ADO58854 Human MUC
19	109	100.0	25	3	ADC78317 MUC1 pept
20	109	100.0	25	6	ABU07528 Human N-a
21	109	100.0	25	7	AD25858 N-acetyl
22	109	100.0	25	8	ADF32636 MUC-1 rel
23	109	100.0	25	8	ADO58858 Human MUC
24	109	100.0	28	2	AAW03361 Mucin tan

25	109	100.0	30	8	ADK65927	Adk65927 Anti-muri
26	109	100.0	31	4	AAW46087	Aaw46087 Human MUC
27	109	100.0	40	2	AAR68002	Aar68002 Mucin rep
28	109	100.0	40	2	AAW54873	Aaw54873 Carcinoma
29	109	100.0	40	2	AAW72703	Aaw72703 Human muc
30	109	100.0	40	5	ABP56039	Abp56039 Repeat mo
31	109	100.0	40	8	ADF32639	Adf32639 MUC-1 rel
32	109	100.0	40	9	ADK05275	Adk05275 Human muc
33	109	100.0	40	9	ADY53503	Ady53503 MUC1 grow
34	109	100.0	41	3	AAW96170	Aaw96170 MUC1 repe
35	109	100.0	43	6	AEE33952	Aee33952 Dilipidat
36	109	100.0	43	7	ADD88870	Add88870 Synthetic
37	109	100.0	46	5	ABP56035	Abp56035 Histidine
38	109	100.0	46	6	AAE33936	Aae33936 MUC 1 pep
39	109	100.0	46	9	ADY53499	Ady53499 His-tagge
40	109	100.0	50	2	AAW35739	Aaw35739 Mucin pep
41	109	100.0	51	2	AAW31697	Aaw31697 Mucin pep
42	109	100.0	100	5	ABB76181	Abb76181 Synthetic
43	109	100.0	105	2	AAR68022	Aar68022 Mucin pep
44	109	100.0	105	2	AAW72697	Aaw72697 Synthetic
45	109	100.0	109	8	ADI57744	Adi57744 Human bre

ALIGNMENTS

RESULT 1

AAW25112

ID AAY25112 standard; peptide; 20 AA.

XX AC AAY25112;

XX DT 25-AUG-1999 (first entry)

XX DE Synthetic glycopeptide #2 homologous to MUC1.

XX KW Immunodominant; epitope; MUC1; epithelial mucin; tumour; vaccine; mammary carcinoma; colorectal carcinoma; pancreas carcinoma.

XX OS Synthetic.

XX PN DE19758400-A1.

XX PD 01-JUL-1999.

XX PF 30-DEC-1997; 97DE-01058400.

XX PR 30-DEC-1997; 97DE-01058400.

XX PA (DELB-) DELBRUECK CENT MOLEKULARE MEDIZIN MAX. (HANI/) HANISCH F.

XX PI Karsten U, Hanisch F, Paulsen H;

XX DR WPI; 1999-372374/32.

XX PT New tumour vaccine - for treating tumour cells of mammary tumours, colorectal and pancreas carcinomas.

XX PS Example 1; Page 3; 6pp; German.

XX CC This invention describes a novel tumour vaccine which comprises a synthetic peptide derived from human epithelial mucin MUC1. This active tumour vaccine can be used against tumour cells from mammary, colorectal or pancreas carcinomas. This sequence represents a synthetic glycopeptide which is homologous to an immunodominant epitope from human MUC1

XX SQ Sequence 20 AA;

Query Match 100.0%; Score 109; DB 2; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.2e-06;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HGVTAPDTRPAPGSTAPPA 20
 |||||
 Db 1 HGVTAPDTRPAPGSTAPPA 20

RESULT 4
 AEB96326
 ID AEB96326 standard; peptide; 20 AA.
 XX
 AC AEB96326;
 XX
 DT 06-OCT-2005 (first entry)
 XX
 DE MUC1 O-linked glycosylation region #1.
 XX
 KW antianemic; nephrotropic; neuroprotective; nootropic; cerebroprotective;
 anti-HIV; anti-inflammatory; virucide; hepatotropic; cytostatic;
 immunosuppressive; respiratory-gen.; muscular-gen.; immunomodulator;
 MUC1.
 XX
 OS Unidentified.
 XX
 PN WO2005070138-A2.
 XX
 PD 04-AUG-2005.
 XX
 PF 10-JAN-2005; 2005WO-US000799.
 XX
 PR 08-JAN-2004; 2004US-0535284P.
 PR 12-FEB-2004; 2004US-0544411P.
 PR 20-FEB-2004; 2004US-0546631P.
 PR 23-MAR-2004; 2004US-0555813P.
 PR 12-MAY-2004; 2004US-0570891P.
 XX
 PA (NEOS-) NEOSE TECHNOLOGIES INC.
 XX
 PI Defrees S, Zopf DA, Wang Z, Clausen H;
 XX
 DR WPI; 2005-597714/51.
 XX
 PT Novel mutant polypeptide having O-linked glycosylation site that does not
 exist in wild-type polypeptide, useful for providing granulocyte colony
 stimulating factor therapy, hormone therapy and interferon therapy.
 XX
 PS Disclosure; Page 32; 197pp; English.
 XX

The invention relates to an isolated polypeptide (I) comprising a mutant
 peptide sequence, where the mutant peptide sequence encodes an O-linked
 glycosylation site that does not exist in a wild-type polypeptide. Also
 disclosed are pharmaceutical compositions (PC1-4). (I) and PC1 are useful
 for providing G-CSF therapy to a subject in need of the therapy. (I) and
 PC2 are useful for providing growth hormone therapy to a subject in need
 of therapy. (I) and PC3 are useful for providing G-CSF therapy to a subject
 in need of therapy. (I) and PC4 are useful for providing interferon
 therapy to a subject in need of therapy. (I) and PC1-PC4 are useful for
 treating general anemia, chronic renal failure, nephritis, and
 thalassemia; neurological disorders such as brain/spine injury,
 Alzheimer's disease and multiple sclerosis, treating AIDS and hepatitis B
 or C, viral infectious caused by a variety of viruses such as human
 papilloma virus (HPV), coronavirus, HIV; cancers such as AIDS-related
 Kaposi's sarcoma, malignant melanoma, renal cancer, bone cancers,
 treating an assortment of other diseases and conditions such as Sjogren's
 syndrome (autoimmune disease), chronic fatigue syndrome, and pulmonary
 fibrosis, treating CNS disorders such as herpes simplex virus (HSV),
 musculoskeletal infections. The present sequence represents the amino
 acid sequence of a MUC1 O-linked glycosylation region.

Sequence 20 AA;
 Query Match 100.0%; Score 109; DB 9; Length 20;
 Best Local Similarity 100.0%; Pred. No. 1.2e-06;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HGVTAPDTRPAPGSTAPPA 20
 |||||
 Db 1 HGVTAPDTRPAPGSTAPPA 20

RESULT 5
 AAY25111
 ID AAY25111 standard; peptide; 21 AA.
 XX
 AC AAY25111;
 XX
 DT 25-AUG-1999 (first entry)
 XX
 DE Synthetic glycopeptide #1 homologous to MUC1.
 XX
 KW Immunodominant; epitope; MUC1; epithelial mucin; tumour; vaccine;
 mammary carcinoma; colorectal carcinoma; pancreas carcinoma.
 XX
 OS Synthetic.
 XX
 PN DE19758400-A1.
 XX
 PD 01-JUL-1999.
 XX
 PF 30-DEC-1997; 97DE-01058400.
 XX
 PR 30-DEC-1997; 97DE-01058400.
 XX
 PA (DELB-) DELBRUECK CENT MOLEKULARE MEDIZIN MAX.
 (HANI/) HANISCH F.
 XX
 PI Karsten U, Hanisch F, Paulsen H;
 XX
 DR WPI; 1999-372374/32.
 XX
 PT New tumour vaccine - for treating tumour cells of mammary tumours,
 colorectal and pancreas carcinomas.
 XX
 PS Example 1; Page 3; 6pp; German.
 XX

This invention describes a novel tumour vaccine which comprises a
 synthetic peptide derived from human epithelial mucin MUC1. This active
 tumour vaccine can be used against tumour cells from mammary, colorectal
 or pancreas carcinomas. This sequence represents a synthetic glycopeptide
 which is homologous to an immunodominant epitope from human MUC1

Sequence 21 AA;
 Query Match 100.0%; Score 109; DB 2; Length 21;
 Best Local Similarity 100.0%; Pred. No. 1.2e-06;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HGVTAPDTRPAPGSTAPPA 20
 |||||
 Db 2 HGVTAPDTRPAPGSTAPPA 21

RESULT 6
 AAY96171
 ID AAY96171 standard; peptide; 21 AA.
 XX
 AC AAY96171;
 XX
 DT 19-DEC-2000 (first entry)
 XX
 DE MUC1 repeat sequence fragment A1.
 XX
 KW MUC1; immunomodulator; glycopeptide; T-lymphocyte; T-cell; proliferation;
 cancer; sarcoma; carcinoma; leukaemia; diagnosis; therapy; vaccine;
 adjuvant; glycosylation.
 XX

OS Homo sapiens.
 XX Key Location/Qualifiers
 FT Modified-site 5
 FT /note= "O-glycosylated by GalNAc-beta-1-3Gal"
 XX
 XX
 PN WO200052046-A1.
 XX
 XX 08-SEP-2000.
 XX
 XX 01-MAR-2000; 2000WO-GB000724.
 XX
 XX 01-MAR-1999; 99GB-00004695.
 XX
 XX (IMCR) IMPERIAL CANCER RES TECHNOLOGY LTD.
 XX
 XX Burchell J, Taylor-Papadimitriou J;
 XX WPI; 2000-601868/57.
 DR
 XX New immunomodulating glycopeptide that causes super-proliferation of T
 XX cells, useful for treating cells in vitro, for diagnosing or treating
 PT cancer (e.g. carcinoma or sarcoma) or as an adjuvant.
 PT
 XX Disclosure; Page 23; 35pp; English.
 PS
 XX The present sequence comprises glycopeptide A1, a fragment of the repeat
 CC sequence of MUC1 glycosylated at Thr-5. Glycopeptides comprising this
 CC fragment of MUC1 repeat sequence, especially having a Gal-GalNAc or
 CC GalNAc moiety on Thr-10 or Thr-17 (see AAY96172-74), are useful as
 CC immunomodulators, causing super-proliferation of T cells. Such
 CC glycopeptides can be used in the treatment or diagnosis of a disease, in
 CC particular cancer, or as vaccine adjuvants. The glycopeptides are
 CC particularly useful in manufacturing a medicament for preventing or
 CC treating cancer by stimulating T cells whose receptors recognize the
 CC glycopeptide. They are also useful for diagnosing or treating cancer,
 CC e.g. carcinoma (e.g. mammary, lung, bladder or colon carcinomas, or ovary
 CC and endometrial tumours), or sarcoma (e.g. soft tissue and bone sarcomas,
 CC or leukaemia). Human peripheral blood lymphocytes (PBL) were used in a
 CC proliferation assay. The proliferation index of A1 (taking the index as 1
 CC when no glycopeptide was present) was 0.7-1.9
 XX
 SQ Sequence 21 AA;
 Query Match 100.0%; Score 109; DB 3; Length 21;
 Best Local Similarity 100.0%; Pred. No. 1.2e-06;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 HGVTSAPDTRPAPGSTAPPA 20
 Db 2 HGVTSAPDTRPAPGSTAPPA 21
 RESULT 7
 AAY96173
 ID AAY96173 standard; peptide; 21 AA.
 XX
 AC AAY96173;
 XX
 DT 19-DEC-2000 (first entry)
 XX
 XX MUC1 repeat sequence fragment A2.
 DE
 XX MUC1; immunomodulator; glycopeptide; T-lymphocyte; T-cell; proliferation;
 KW cancer; sarcoma; carcinoma; leukaemia; diagnosis; therapy; vaccine;
 KW adjuvant; glycosylation.
 XX
 XX Homo sapiens.
 OS
 XX Key Location/Qualifiers
 FT Modified-site 10
 FT /note= "O-glycosylated by GalNAc-beta-1-3Gal"
 FT
 FT
 XX

PN WO200052046-A1.
 XX
 PD 08-SEP-2000.
 XX
 PF 01-MAR-2000; 2000WO-GB000724.
 XX
 XX 01-MAR-1999; 99GB-00004695.
 XX
 XX (IMCR) IMPERIAL CANCER RES TECHNOLOGY LTD.
 XX
 XX Burchell J, Taylor-Papadimitriou J;
 XX WPI; 2000-601868/57.
 DR
 XX New immunomodulating glycopeptide that causes super-proliferation of T
 XX cells, useful for treating cells in vitro, for diagnosing or treating
 PT cancer (e.g. carcinoma or sarcoma) or as an adjuvant.
 PT
 XX Disclosure; Page 23; 35pp; English.
 PS
 XX The present sequence comprises glycopeptide A2, a fragment of the repeat
 CC sequence of MUC1 glycosylated at Thr-10. Glycopeptides comprising this
 CC fragment of MUC1 repeat sequence, especially having a Gal-GalNAc or
 CC GalNAc moiety on Thr-10 or Thr-17 (see AAY96172-74), are useful as
 CC immunomodulators, causing super-proliferation of T cells. Such
 CC glycopeptides can be used in the treatment or diagnosis of a disease, in
 CC particular cancer, or as vaccine adjuvants. The glycopeptides are
 CC particularly useful in manufacturing a medicament for preventing or
 CC treating cancer by stimulating T cells whose receptors recognize the
 CC glycopeptide. They are also useful for diagnosing or treating cancer,
 CC e.g. carcinoma (e.g. mammary, lung, bladder or colon carcinomas, or ovary
 CC and endometrial tumours), or sarcoma (e.g. soft tissue and bone sarcomas,
 CC or leukaemia). Human peripheral blood lymphocytes (PBL) were used in a
 CC proliferation assay. The proliferation index of A2 (taking the index as 1
 CC when no glycopeptide was present) was up to 55
 XX
 SQ Sequence 21 AA;
 Query Match 100.0%; Score 109; DB 3; Length 21;
 Best Local Similarity 100.0%; Pred. No. 1.2e-06;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 HGVTSAPDTRPAPGSTAPPA 20
 Db 2 HGVTSAPDTRPAPGSTAPPA 21
 RESULT 8
 AAY96169
 ID AAY96169 standard; peptide; 21 AA.
 XX
 AC AAY96169;
 XX
 DT 19-DEC-2000 (first entry)
 XX
 XX MUC1 repeat sequence fragment.
 DE
 XX MUC1; immunomodulator; glycopeptide; T-lymphocyte; T-cell; proliferation;
 KW cancer; sarcoma; carcinoma; leukaemia; diagnosis; therapy; vaccine;
 KW adjuvant; glycosylation.
 XX
 XX Homo sapiens.
 OS
 XX WO200052046-A1.
 XX
 XX 08-SEP-2000.
 XX
 XX 01-MAR-2000; 2000WO-GB000724.
 XX
 XX 01-MAR-1999; 99GB-00004695.
 XX
 XX (IMCR) IMPERIAL CANCER RES TECHNOLOGY LTD.
 XX
 XX

PI Burchell J, Taylor-Papadimitriou J;
 XX WPI; 2000-601868/57.
 XX
 XX New immunomodulating glycopeptide that causes super-proliferation of T
 PT cells, useful for treating cells in vitro, for diagnosing or treating
 PT cancer (e.g. carcinoma or sarcoma) or as an adjuvant.
 XX
 XX Claim 2; Page 3; 35pp; English.
 XX
 XX The present sequence comprises a fragment of the repeat sequence of MUC1.
 CC A glycopeptide comprising this sequence, especially having a Gal-GalNAC
 CC or GalNAC moiety on Thr-10 or Thr-17 (see AAY96172-74), is useful as an
 CC immunomodulator causing super-proliferation of T cells. Such
 CC glycopeptides can be used in the treatment or diagnosis of a disease, in
 CC particular cancer, or as a vaccine adjuvant. The glycopeptides are
 CC particularly useful in manufacturing a medicament for preventing or
 CC treating cancer by stimulating T cells whose receptors recognize the
 CC glycopeptide. They are also useful for diagnosing or treating cancer,
 CC e.g. carcinoma (e.g. mammary, lung, bladder or colon carcinomas, or ovary
 CC and endometrial tumours), or sarcoma (e.g. soft tissue and bone sarcomas,
 CC or leukaemia)
 XX
 XX Sequence 21 AA;
 SQ
 Query Match 100.0%; Score 109; DB 3; Length 21;
 Best Local Similarity 100.0%; Pred. No. 1.2e-06;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 HGVTSA PDTRPAGSTAPPA 20
 Db 2 HGVTSA PDTRPAGSTAPPA 21
 |||||
 RESULT 9
 AAY96180
 ID AAY96180 standard; peptide; 21 AA.
 XX
 AC AAY96180;
 XX
 DT 19-DEC-2000 (first entry)
 XX
 DE MUC1 repeat sequence fragment A9.
 XX
 XX MUC1; immunomodulator; glycopeptide; T-lymphocyte; T-cell; proliferation;
 KW cancer; sarcoma; carcinoma; leukaemia; diagnosis; therapy; vaccine;
 KW adjuvant; glycosylation.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FH Modified-site 5
 FT Modified-site 6 /note= "O-glycosylated by GalNac-beta-1-3Gal"
 FT Modified-site 16 /note= "O-glycosylated by GalNac-beta-1-3Gal"
 FT Modified-site 17 /note= "O-glycosylated by GalNac-beta-1-3Gal"
 FT Modified-site 17 /note= "O-glycosylated by GalNac-beta-1-3Gal"
 XX
 PN WO200052046-A1.
 XX
 PD 08-SEP-2000.
 XX
 XX 01-MAR-2000; 2000WO-GB0000724.
 XX
 XX 01-MAR-1999; 99GB-00004695.
 XX
 XX (IMCR) IMPERIAL CANCER RES TECHNOLOGY LTD.
 XX
 XX Burchell J, Taylor-Papadimitriou J;
 XX WPI; 2000-601868/57.

XX New immunomodulating glycopeptide that causes super-proliferation of T
 PT cells, useful for treating cells in vitro, for diagnosing or treating
 PT cancer (e.g. carcinoma or sarcoma) or as an adjuvant.
 XX
 PS Disclosure; Page 23; 35pp; English.
 XX
 CC The present sequence comprises glycopeptide A9, a fragment of the repeat
 CC sequence of MUC1 glycosylated at Thr-5, Ser-6, Thr-16 and Thr-17.
 CC Glycopeptides comprising this fragment of MUC1 repeat sequence,
 CC especially having a Gal-GalNAC or GalNAC moiety on Thr-10 or Thr-17 (see
 CC AAY96172-74), are useful as immunomodulators, causing super-proliferation
 CC of T cells. Such glycopeptides can be used in the treatment or diagnosis
 CC of a disease, in particular cancer, or as vaccine adjuvants. The
 CC glycopeptides are particularly useful in manufacturing a medicament for
 CC preventing or treating cancer by stimulating T cells whose receptors
 CC recognize the glycopeptide. They are also useful for diagnosing or
 CC treating cancer, e.g. carcinoma (e.g. mammary, lung, bladder or colon
 CC carcinomas, or ovary and endometrial tumours), or sarcoma (e.g. soft
 CC tissue and bone sarcomas, or leukaemia). Human peripheral blood
 CC lymphocytes were used in a proliferation assay. The proliferation index
 CC of A9 (taking the index as 1 when no glycopeptide was present) was up to
 CC 34
 XX
 XX Sequence 21 AA;
 SQ
 Query Match 100.0%; Score 109; DB 3; Length 21;
 Best Local Similarity 100.0%; Pred. No. 1.2e-06;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 HGVTSA PDTRPAGSTAPPA 20
 Db 2 HGVTSA PDTRPAGSTAPPA 21
 |||||
 RESULT 10
 AAY96179
 ID AAY96179 standard; peptide; 21 AA.
 XX
 AC AAY96179;
 XX
 DT 19-DEC-2000 (first entry)
 XX
 DE MUC1 repeat sequence fragment A8.
 XX
 XX MUC1; immunomodulator; glycopeptide; T-lymphocyte; T-cell; proliferation;
 KW cancer; sarcoma; carcinoma; leukaemia; diagnosis; therapy; vaccine;
 KW adjuvant; glycosylation.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FH Modified-site 5
 FT Modified-site 6 /note= "O-glycosylated by GalNac-beta-1-3Gal"
 FT Modified-site 10 /note= "O-glycosylated by GalNac-beta-1-3Gal"
 FT Modified-site 16 /note= "O-glycosylated by GalNac-beta-1-3Gal"
 FT Modified-site 17 /note= "O-glycosylated by GalNac-beta-1-3Gal"
 FT Modified-site 17 /note= "O-glycosylated by GalNac-beta-1-3Gal"
 XX
 PN WO200052046-A1.
 XX
 PD 08-SEP-2000.
 XX
 XX 01-MAR-2000; 2000WO-GB0000724.
 XX
 XX 01-MAR-1999; 99GB-00004695.
 XX
 XX (IMCR) IMPERIAL CANCER RES TECHNOLOGY LTD.

```

PI Burchell J, Taylor-Papadimitriou J;
XX WPI; 2000-601868/57.
DR
XX
XX New immunomodulating glycopeptide that causes super-proliferation of T
PT cells, useful for treating cells in vitro, for diagnosing or treating
PT cancer (e.g. carcinoma or sarcoma) or as an adjuvant.
XX
XX Disclosure; Page 23; 35pp; English.
PS
XX
XX The present sequence comprises glycopeptide A8, a fragment of the repeat
CC sequence of MUC1 glycosylated at Thr-5, Ser-6, Thr-10, Ser-16 and Thr-17.
CC Glycopeptides comprising this fragment of MUC1 repeat sequence,
CC especially having a Gal-GalNac or GalNac moiety on Thr-10 or Thr-17 (see
CC AAY96172-74), are useful as immunomodulators, causing super-proliferation
CC of T cells. Such glycopeptides can be used in the treatment or diagnosis
CC of a disease, in particular cancer, or as vaccine adjuvants. The
CC glycopeptides are particularly useful in manufacturing a medicament for
CC preventing or treating cancer by stimulating T cells whose receptors
CC recognize the glycopeptide. They are also useful for diagnosing or
CC treating cancer, e.g. carcinoma (e.g. mammary, lung, bladder or colon
CC carcinomas, or ovary and endometrial tumours), or sarcoma (e.g. soft
CC tissue and bone sarcomas, or leukaemia). Human peripheral blood
CC lymphocytes were used in a proliferation assay. The proliferation index
CC of A8 (taking the index as 1 when no glycopeptide was present) was up to
CC 7
XX
XX Sequence 21 AA;
SQ
Query Match 100.0%; Score 109; DB 3; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.2e-06;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HGVTSPDTRPAPGSTAPPA 20
DB 2 HGVTSPDTRPAPGSTAPPA 21
|||||
|||||

RESULT 11
AAY96172
ID AAY96172 standard; peptide; 21 AA.
AC AAY96172;
XX
XX 19-DEC-2000 (first entry)
DT
XX
XX MUC1 repeat sequence fragment A2 GalNac.
DE
XX
XX MUC1; immunomodulator; glycopeptide; T-lymphocyte; T-cell; proliferation;
KW cancer; sarcoma; carcinoma; leukaemia; diagnosis; therapy; vaccine;
KW adjuvant; glycosylation.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH Modified-site 10
FT /note= "O-glycosylated by GalNac"
FT
XX
XX WO200052046-A1.
FN
XX
XX 08-SEP-2000.
PD
XX
XX 01-MAR-2000; 2000WO-GB0000724.
PF
XX
XX 01-MAR-1999; 99GB-00004695.
PR
XX
XX (IMCR ) IMPERIAL CANCER RES TECHNOLOGY LTD.
PA
XX Burchell J, Taylor-Papadimitriou J;
PI WPI; 2000-601868/57.
XX
XX New immunomodulating glycopeptide that causes super-proliferation of T
PT cells, useful for treating cells in vitro, for diagnosing or treating
PT cancer (e.g. carcinoma or sarcoma) or as an adjuvant.
XX
XX Disclosure; Page 23; 35pp; English.
PS
XX
XX The present sequence comprises glycopeptide A3, a fragment of the repeat
CC

```

cells, useful for treating cells in vitro, for diagnosing or treating cancer (e.g. carcinoma or sarcoma) or as an adjuvant.

Disclosure; Page 23; 35pp; English.

The present sequence comprises glycopeptide A2 GalNac, a fragment of the repeat sequence of MUC1 glycosylated at Thr-10. Glycopeptides comprising this fragment of MUC1 repeat sequence, especially having a Gal-GalNac or GalNac moiety on Thr-10 or Thr-17 (see AAY96172-74), are useful as immunomodulators, causing super-proliferation of T cells. Such glycopeptides can be used in the treatment or diagnosis of a disease, in particular cancer, or as vaccine adjuvants. The glycopeptides are particularly useful in manufacturing a medicament for preventing or treating cancer by stimulating T cells whose receptors recognize the glycopeptide. They are also useful for diagnosing or treating cancer, e.g. carcinoma (e.g. mammary, lung, bladder or colon carcinomas, or ovary and endometrial tumours), or sarcoma (e.g. soft tissue and bone sarcomas, or leukaemia). Human peripheral blood lymphocytes (PBL) were used in a proliferation assay. The proliferation index of A2 GalNac (taking the index as 1 when no glycopeptide was present) was 48

Sequence 21 AA;

Query Match 100.0%; Score 109; DB 3; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.2e-06;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HGVTSPDTRPAPGSTAPPA 20
DB 2 HGVTSPDTRPAPGSTAPPA 21
|||||
|||||

RESULT 12
AAY96174
ID AAY96174 standard; peptide; 21 AA.

AC AAY96174;

XX 19-DEC-2000 (first entry)

XX MUC1 repeat sequence fragment A3.

MUC1; immunomodulator; glycopeptide; T-lymphocyte; T-cell; proliferation; cancer; sarcoma; carcinoma; leukaemia; diagnosis; therapy; vaccine; adjuvant; glycosylation.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Modified-site 17 /note= "O-glycosylated by GalNac-beta-1-3Gal"

XX WO200052046-A1.

XX 08-SEP-2000.

XX 01-MAR-2000; 2000WO-GB0000724.

XX 01-MAR-1999; 99GB-00004695.

XX (IMCR) IMPERIAL CANCER RES TECHNOLOGY LTD.

XX Burchell J, Taylor-Papadimitriou J;

XX WPI; 2000-601868/57.

XX New immunomodulating glycopeptide that causes super-proliferation of T cells, useful for treating cells in vitro, for diagnosing or treating cancer (e.g. carcinoma or sarcoma) or as an adjuvant.

XX Disclosure; Page 23; 35pp; English.

XX The present sequence comprises glycopeptide A3, a fragment of the repeat

CC sequence of MUC1 glycosylated at Thr-17. Glycopeptides comprising this
 CC fragment of MUC1 repeat sequence, especially having a Gal-GalNAc or
 CC GalNAc moiety on Thr-10 or Thr-17 (see AAY96172-74), are useful as
 CC immunomodulators, causing super-proliferation of T cells. Such
 CC glycopeptides can be used in the treatment or diagnosis of a disease, in
 CC particular cancer, or as vaccine adjuvants. The glycopeptides are
 CC particularly useful in manufacturing a medicament for preventing or
 CC treating cancer by stimulating T cells whose receptors recognize the
 CC glycopeptide. They are also useful for diagnosing or treating cancer,
 CC e.g. carcinoma (e.g. mammary, lung, bladder or colon carcinomas, or ovary
 CC and endometrial tumours), or sarcoma (e.g. soft tissue and bone sarcomas,
 CC or leukaemia). Human peripheral blood lymphocytes (PBL) were used in a
 CC proliferation assay. The proliferation index of A3 (taking the index as 1
 CC when no glycopeptide was present) was up to 87
 XX
 SQ Sequence 21 AA;

Query Match 100.0%; Score 109; DB 3; Length 21;
 Best Local Similarity 100.0%; Pred. No. 1.2e-06;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HGVTSAPDTRPAPGSTAPPA 20
 |||||
 DB 2 HGVTSAPDTRPAPGSTAPPA 21
 |||||

RESULT 13
 AAY96176
 ID AAY96176 standard; peptide; 21 AA.
 XX
 AC AAY96176;
 XX

19-DEC-2000 (first entry)

MUC1 repeat sequence fragment A5.

MUC1; immunomodulator; glycopeptide; T-lymphocyte; T-cell; proliferation;
 cancer; sarcoma; carcinoma; leukaemia; diagnosis; therapy; vaccine;
 adjuvant; glycosylation.
 Homo sapiens.

Key Location/Qualifiers
 FT Modified-site 16 /note= "O-glycosylated by GalNAc-beta-1-3Gal"
 FT

WO200052046-A1.

08-SEP-2000.

01-MAR-2000; 2000WO-GB000724.

01-MAR-1999; 99GB-00004695.

(IMCR) IMPERIAL CANCER RES TECHNOLOGY LTD.

Burchell J, Taylor-Papadimitriou J;

WPI; 2000-601868/57.

New immunomodulating glycopeptide that causes super-proliferation of T
 cells, useful for treating cells in vitro, for diagnosing or treating
 cancer (e.g. carcinoma or sarcoma) or as an adjuvant.

Disclosure; Page 23; 35pp; English.

The present sequence comprises glycopeptide A5, a fragment of the repeat
 sequence of MUC1 glycosylated at Ser-16. Glycopeptides comprising this
 fragment of MUC1 repeat sequence, especially having a Gal-GalNAc or
 CC GalNAc moiety on Thr-10 or Thr-17 (see AAY96172-74), are useful as
 CC immunomodulators, causing super-proliferation of T cells. Such
 CC glycopeptides can be used in the treatment or diagnosis of a disease, in
 CC particular cancer, or as vaccine adjuvants. The glycopeptides are

CC particularly useful in manufacturing a medicament for preventing or
 CC treating cancer by stimulating T cells whose receptors recognize the
 CC glycopeptide. They are also useful for diagnosing or treating cancer,
 CC e.g. carcinoma (e.g. mammary, lung, bladder or colon carcinomas, or ovary
 CC and endometrial tumours), or sarcoma (e.g. soft tissue and bone sarcomas,
 CC or leukaemia). Human peripheral blood lymphocytes (PBL) were used in a
 CC proliferation assay. The proliferation index of A5 (taking the index as 1
 CC when no glycopeptide was present) was up to 31
 XX
 SQ Sequence 21 AA;

Query Match 100.0%; Score 109; DB 3; Length 21;
 Best Local Similarity 100.0%; Pred. No. 1.2e-06;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HGVTSAPDTRPAPGSTAPPA 20
 |||||
 DB 2 HGVTSAPDTRPAPGSTAPPA 21
 |||||

RESULT 14
 AAY96178
 ID AAY96178 standard; peptide; 21 AA.
 XX
 AC AAY96178;
 XX

19-DEC-2000 (first entry)

MUC1 repeat sequence fragment A7.

MUC1; immunomodulator; glycopeptide; T-lymphocyte; T-cell; proliferation;
 cancer; sarcoma; carcinoma; leukaemia; diagnosis; therapy; vaccine;
 adjuvant; glycosylation.

Homo sapiens.

Key Location/Qualifiers
 FT Modified-site 5 /note= "O-glycosylated by GalNAc-beta-1-3Gal"
 FT Modified-site 16 /note= "O-glycosylated by GalNAc-beta-1-3Gal"
 FT Modified-site 17 /note= "O-glycosylated by GalNAc-beta-1-3Gal"
 FT

WO200052046-A1.

08-SEP-2000.

01-MAR-2000; 2000WO-GB000724.

01-MAR-1999; 99GB-00004695.

(IMCR) IMPERIAL CANCER RES TECHNOLOGY LTD.

Burchell J, Taylor-Papadimitriou J;

WPI; 2000-601868/57.

New immunomodulating glycopeptide that causes super-proliferation of T
 cells, useful for treating cells in vitro, for diagnosing or treating
 cancer (e.g. carcinoma or sarcoma) or as an adjuvant.

Disclosure; Page 23; 35pp; English.

The present sequence comprises glycopeptide A7, a fragment of the repeat
 sequence of MUC1 glycosylated at Thr-5, Ser-16 and Thr-17. Glycopeptides
 comprising this fragment of MUC1 repeat sequence, especially having a Gal
 -GalNAc or GalNAc moiety on Thr-10 or Thr-17 (see AAY96172-74), are
 useful as immunomodulators, causing super-proliferation of T cells. Such
 glycopeptides can be used in the treatment or diagnosis of a disease, in
 particular cancer, or as vaccine adjuvants. The glycopeptides are
 particularly useful in manufacturing a medicament for preventing or
 treating cancer by stimulating T cells whose receptors recognize the

CC proliferation assay. The proliferation index of A6 (taking the index as 1
CC when no glycopeptide was present) was up to 34
XX
SQ Sequence 21 AA;
Query Match 100.0%; Score 109; DB 3; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.2e-06;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 HGVTSAPDTRPAPGSTAPPA 20
Db 2 HGVTSAPDTRPAPGSTAPPA 21
Search completed: April 14, 2006, 10:24:15
Job time : 298.573 secs

glycopeptide. They are also useful for diagnosing or treating cancer,
e.g. carcinoma (e.g. mammary, lung, bladder or colon carcinomas, or ovary
and endometrial tumours), or sarcoma (e.g. soft tissue and bone sarcomas,
or leukaemia). Human peripheral blood lymphocytes were used in a
proliferation assay. The proliferation index of A7 (taking the index as 1
when no glycopeptide was present) was up to 13
XX
SQ Sequence 21 AA;
Query Match 100.0%; Score 109; DB 3; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.2e-06;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 HGVTSAPDTRPAPGSTAPPA 20
Db 2 HGVTSAPDTRPAPGSTAPPA 21

RESULT 15
AAY96177
ID AAY96177 standard; peptide; 21 AA.
XX
AC AAY96177;
XX
DT 19-DEC-2000 (first entry)
XX
DE MUC1 repeat sequence fragment A6.
XX
KW MUC1; immunomodulator; glycopeptide; T-lymphocyte; T-cell; proliferation;
cancer; sarcoma; carcinoma; leukaemia; diagnosis; therapy; vaccine;
adjuvant; glycosylation.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Modified-site 5 /note= "O-glycosylated by GalNAc-beta-1-3Gal"
FT Modified-site 17 /note= "O-glycosylated by GalNAc-beta-1-3Gal"
FT
XX WO200052046-A1.
XX
XX 08-SEP-2000.
XX
XX 01-MAR-2000; 2000WO-GB000724.
XX
XX 01-MAR-1999; 99GB-00004695.
XX (IMCR) IMPERIAL CANCER RES TECHNOLOGY LTD.
XX
XX Burchell J, Taylor-Papadimitriou J;
XX WPI; 2000-601868/57.
XX
XX New immunomodulating glycopeptide that causes super-proliferation of T
cells, useful for treating cells in vitro, for diagnosing or treating
cancer (e.g. carcinoma or sarcoma) or as an adjuvant.
XX
XX Disclosure; Page 23; 35pp; English.
XX
XX The present sequence comprises glycopeptide A6, a fragment of the repeat
sequence of MUC1 glycosylated at Thr-5 and Thr-17. Glycopeptides
comprising this fragment of MUC1 repeat sequence, especially having a Gal
-GalNAc or GalNAc moiety on Thr-10 or Thr-17 (see AAY96172-74), are
useful as immunomodulators, causing super-proliferation of T cells. Such
glycopeptides can be used in the treatment or diagnosis of a disease, in
particular cancer, or as vaccine adjuvants. The glycopeptides are
particularly useful in manufacturing T cells whose receptors recognize the
treating cancer by stimulating T cells whose receptors recognize the
glycopeptide. They are also useful for diagnosing or treating cancer,
e.g. carcinoma (e.g. mammary, lung, bladder or colon carcinomas, or ovary
and endometrial tumours), or sarcoma (e.g. soft tissue and bone sarcomas,
or leukaemia). Human peripheral blood lymphocytes were used in a

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2006, 10:24:48 ; Search time 23.4146 Seconds
(without alignments)
82.185 Million cell updates/sec

Title: US-09-606-910E-2

Perfect score: 109
Sequence: 1 HGVTSAPDTRPAPGSTAPPA 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 80.*

- 1: pir1.*
- 2: pir2.*
- 3: pir3.*
- 4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	109	100.0	347	1 S10571	mucin 1 precursor,
2	109	100.0	1344	2 A35175	mucin 1 precursor,
3	93	85.3	256	2 A60533	tumor-associated a
4	62	56.9	1334	2 T50568	probable multi-dom
5	57	52.3	428	2 T24769	hypothetical prote
6	56	51.4	267	2 F87665	hypothetical prote
7	56	51.4	569	2 A13347	hypothetical prote
8	56	51.4	2706	2 T28155	variant-specific s
9	55	50.5	4391	2 A38096	perlecan precursor
10	53	48.6	373	2 A70856	probable lppz prot
11	53	48.6	417	2 A54416	prostatcyclin recep
12	53	48.6	708	2 JC4364	gelatinase B (EC 3
13	53	48.6	708	2 S62907	gelatinase B (EC 3
14	52	47.7	115	2 F72570	hypothetical prote
15	52	47.7	182	2 S73046	penicillin binding
16	52	47.7	196	2 T51838	blue copper bindin
17	52	47.7	411	2 D86995	probable D-alanyl-
18	51.5	47.2	635	2 F75477	hypothetical prote
19	51	46.8	217	2 A31887	hypothetical prote
20	51	46.8	257	2 T10586	small nuclear ribo
21	51	46.8	310	2 A70671	probable puru prot
22	51	46.8	382	2 T14336	RAD23 protein, iso
23	51	46.8	1004	2 JC2221	major surface glyc
24	51	46.8	1199	2 JC4816	major surface glyc
25	51	46.8	1201	2 A83007	hypothetical prote
26	50	45.9	402	2 S73773	dihydrolipoamide a
27	50	45.9	571	2 C75530	conserved hypothet
28	50	45.9	729	2 F70803	hypothetical prote
29	50	45.9	730	2 JC1456	gelatinase B (EC 3

30	50	45.9	1173	1 A53430	1-phosphatidylinos
31	50	45.9	1216	2 A28821	1-phosphatidylinos
32	49	45.0	78	2 T27876	hypothetical prote
33	49	45.0	175	2 T27875	hypothetical prote
34	49	45.0	430	2 JC7379	levansucrase (EC 2
35	49	45.0	447	2 T18264	cellulosome anchor
36	49	45.0	594	2 T12995	pectinesterase hom
37	49	45.0	630	2 A39344	tumor-associated m
38	49	45.0	631	2 I52257	episialin - mouse
39	49	45.0	1076	2 T30842	serine-repeat anti
40	49	45.0	1101	2 G70951	probable ATP-depen
41	49	45.0	1255	2 T31065	diaphanous protein
42	49	45.0	1315	2 A56101	collagen alpha 1(X
43	49	45.0	1774	2 B56101	collagen alpha 1(X
44	48.5	44.5	205	2 T34724	probable membrane
45	48.5	44.5	464	2 T36256	probable membrane

ALIGNMENTS

RESULT 1

S10571
mucin 1 precursor, secreted epithelial tumor antigen splice form - human
N:Contains: mucin 1 secreted breast-cancer-associated splice form
C:Species: Homo sapiens (man)
C:Date: 07-Apr-1994 #sequence revision 07-Apr-1994 #text_change 09-Jul-2004
C:Accession: S10571; JN0100; I56024; S09706; S10217
R:Wreschner, D.H.; Hareuveni, M.; Tsefaty, I.; Smorodinsky, N.; Horev, J.; Zaretsky, J.
Eur. J. Biochem. 189, 463-473, 1990
A>Title: Human epithelial tumor antigen cDNA sequences. Differential splicing may gener.
A:Reference number: S10571; MUID:90276413; PMID:2351132
A:Accession: S10571
A:Molecule type: mRNA

A:Residues: 1-347 <WR>
A:Cross-references: UNIPROT:Q9UM18; UNIPARC:UPI000016B013; EMBL:X52228; NID:936434; PIDN:
R:Tsarfaty, I.; Hareuveni, M.; Horev, J.; Zaretsky, J.; Weiss, M.; Jeltsch, J.M.; Garni
J. Xing, P.X.; Tjandra, J.J.; Reynolds, K.; McLaughlin, P.J.; Purcell, D.F.J.; McKenzie,
J. Immunol. 142, 3503-3509, 1989
A>Title: Reactivity of anti-human milk fat globule antibodies with synthetic peptides.
A:Reference number: I56024; MUID:89235154; PMID:2715633
A:Accession: JN0100; MUID:91033045; PMID:1688329

A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-135, 'Q', 137-142, 'E', 144-163, 204-208, 'A', 210-347 <TSA>
A:Cross-references: UNIPARC:UPI000016A888; GB:M35093; NID:9182252; PIDN:AAB59612.1; PID
R:Xing, P.X.; Tjandra, J.J.; Reynolds, K.; McLaughlin, P.J.; Purcell, D.F.J.; McKenzie,
J. Immunol. 142, 3503-3509, 1989
A>Title: Reactivity of anti-human milk fat globule antibodies with synthetic peptides.
A:Reference number: I56024; MUID:89235154; PMID:2715633
A:Accession: I56024

A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 182-201 <RES>
A:Cross-references: UNIPARC:UPI00000330DA; GB:M26316; NID:9516622; PIDN:AAA36336.1; PID
R:rendler, S.J.B.
Biochem. J. 267, 733-737, 1990
A>Title: Elements of secondary structure in a human epithelial mucin core peptide fragm

A:Reference number: S09706; MUID:90253387; PMID:2339983
A:Accession: S09706
A:Molecule type: protein
A:Residues: 182-201 <TEN>
A:Cross-references: UNIPARC:UPI00000330DA
C:Genetics:

A:Gene: GDB:MUC1; PUM
A:Cross-references: GDB:120705; OMIM:158340
A:Map position: Iq21-Iq23
C:Keywords: alternative splicing; tandem repeat
F:1-23/Domain: signal sequence #status predicted <SIG>
F:24-347/Product: mucin 1, secreted epithelial tumor antigen splice form #status predic
F:24-163,204-347/Product: mucin 1, secreted breast-cancer-associated splice form #statu

Query Match 100.0%; Score 109; DB 2; Length 347;
Best Local Similarity 100.0%; Pred. No. 2e-06;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HGVTSAPDTRPAGSTAPPA 20
|||||

Db 176 HGVTSAPDTRPAGSTAPPA 195

RESULT 2

A:Residues: 1-142,'Q',144-162,'Q',164-168 <ABE>
A:Cross-references: UNIPARC:UPI000017456A; EMBL:M31823; NID:g181542; PIDN:AAA35757.1; PI
R:Masuzawa, Y.; Miyauchi, T.; Hamanoue, M.; Ando, S.; Yoshida, J.; Takao, S.; Shimazu, H
J. Biochem. 112, 609-615, 1992

A:Title: A novel core protein as well as polymorphic epithelial mucin carry peanut agglu
A:Reference number: JX0235; MUID:93123189; PMID:1478919

A:Accession: PX0066

A:Molecule type: mRNA

A:Residues: 998-1011,'ES',1014-1017,1018-1032,'T',1034-1037,1038-1057 <MAS>
A:Cross-references: UNIPARC:UPI000017456B; UNIPARC:UPI000017456C; UNIPARC:UPI000017456D
A:Experimental source: Gastric carcinoma cell

A:Zrihan-Licht, S.; Baruch, A.; Elroy-Stein, O.; Keydar, I.; Wreschner, D.H.
FEBS Lett. 356, 130-136, 1994

A:Title: Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins cytokine r
A:Reference number: S51026; MUID:95080414; PMID:7988707

A:Contents: annotation

A:Note: undetermined tyrosine residues in the carboxyl-terminal non-repetitive region ar
C:Comment: This protein is length polymorphic. Individuals may have between 21 and 125 c
partial repeats. The repeat shown is defined by SmaI nuclease sites.

C:Comment: Serine and threonine residues in the tandem repeat domain are extensively gly
C:Comment: For an alternative splice form without a tandem repeat domain, see PIR:S48146

C:Genetics:

A:Gene: GDB:MUC1; PUM

A:Cross-references: GDB:120705; OMIM:158340

A:Map position: 1q21-1q23

A:introns: 20/1; 62/3; 1165/3; 1184/2; 1230/1; 1270/3; 1320/3

C:Superfamily: polymorphic epithelial mucin

C:Keywords: alternative splicing; duplication; glycoprotein; glycoprotein; phosphoprotein; polymorphis
F:1-1344/Product: mucin 1 precursor, splice form A #status predicted <PREA>

F:1-62/Region: mucin 1 amino-terminal non-repetitive

F:1-23/Domain: signal sequence #link PREA #status predicted <SIGA>

F:1-19,29-32/Domain: signal sequence #link PREB #status predicted <SIGB>

F:1-19,29-1344/Product: mucin 1 precursor, splice form B #status predicted <PREB>

F:1-19,29-212,1033-1344/Product: mucin 1 precursor, splice form A #status predicted <PREA>

F:138-1017/Region: 20-residue repeats (GSTAPPAGVTSAPDTRPAP)

F:1143-1344/Region: mucin 1 carboxyl-terminal non-repetitive

F:1245-1272/Domain: transmembrane #status predicted <TRN>

F:1046,1064,1118,1144,1222/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:1213/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 100.0%; Score 109; DB 1; Length 1344;
Best Local Similarity 100.0%; Pred. No. 7.2e-06;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HGVTSAPDTRPAGSTAPPA 20
|||||

Db 145 HGVTSAPDTRPAGSTAPPA 164

RESULT 3

A:Residues: 1-19,29-1109,'S',1111-1339,'A',1341-1344 <LAN>
A:Cross-references: UNIPARC:UPI0000174567; GB:J05582; NID:g189598; PIDN:AAA60019.1; PID:
A:Note: GenBank entry HUMPCAB contains four fewer copies of the tandemly repeated sequ
R:Wreschner, D.H.; Hareuveni, M.; Tsarfaty, I.; Smorodinsky, N.; Horev, J.; Zaretzky, J.
Eur. J. Biochem. 189, 463-473, 1990

A:Title: Human epithelial tumor antigen cDNA sequences. Differential splicing may genera
A:Reference number: S10571; MUID:90276413; PMID:2351132

A:Accession: S10572

A:Molecule type: mRNA

A:Residues: 1-19,29-155,'P',157-175,'P',177-182,'A',184-212,1033-1037,'A',1039-1344 <WRE>
A:Cross-references: UNIPARC:UPI0000174569; EMBL:X52229; NID:g37053

R:Wreschner, D.H.
submitted to the EMBL Data Library, March 1990

A:Reference number: S40293

A:Accession: S40293

A:Molecule type: mRNA

A:Residues: 1-19,29-155,'P',157-175,'P',177-182,'A',184-212,1033-1037,'A',1039-1344 <WRE>
A:Cross-references: UNIPARC:UPI0000160A6; EMBL:X52229; NID:g37053; PIDN:CAA36478.1; PID:
R:Abe, M.; Siddiqui, J.; Kufe, D.
Biochem. Biophys. Res. Commun. 165, 644-649, 1989

A:Title: Sequence analysis of the 5' region of the human DF3 breast carcinoma-associated
A:Reference number: A36735; MUID:90088473; PMID:2597151

A:Accession: A36735

A:Molecule type: mRNA

A:Residues: 1-142,'Q',144-162,'Q',164-168 <ABE>
A:Cross-references: UNIPARC:UPI000017456A; EMBL:M31823; NID:g181542; PIDN:AAA35757.1; PI
R:Masuzawa, Y.; Miyauchi, T.; Hamanoue, M.; Ando, S.; Yoshida, J.; Takao, S.; Shimazu, H
J. Biochem. 112, 609-615, 1992

A:Title: A novel core protein as well as polymorphic epithelial mucin carry peanut agglu
A:Reference number: JX0235; MUID:93123189; PMID:1478919

A:Accession: PX0066

A:Molecule type: mRNA

A:Residues: 998-1011,'ES',1014-1017,1018-1032,'T',1034-1037,1038-1057 <MAS>
A:Cross-references: UNIPARC:UPI000017456B; UNIPARC:UPI000017456C; UNIPARC:UPI000017456D
A:Experimental source: Gastric carcinoma cell

A:Zrihan-Licht, S.; Baruch, A.; Elroy-Stein, O.; Keydar, I.; Wreschner, D.H.
FEBS Lett. 356, 130-136, 1994

A:Title: Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins cytokine r
A:Reference number: S51026; MUID:95080414; PMID:7988707

A:Contents: annotation

A:Note: undetermined tyrosine residues in the carboxyl-terminal non-repetitive region ar
C:Comment: This protein is length polymorphic. Individuals may have between 21 and 125 c
partial repeats. The repeat shown is defined by SmaI nuclease sites.

C:Comment: Serine and threonine residues in the tandem repeat domain are extensively gly
C:Comment: For an alternative splice form without a tandem repeat domain, see PIR:S48146

C:Genetics:

A:Gene: GDB:MUC1; PUM

A:Cross-references: GDB:120705; OMIM:158340

A:Map position: 1q21-1q23

A:introns: 20/1; 62/3; 1165/3; 1184/2; 1230/1; 1270/3; 1320/3

C:Superfamily: polymorphic epithelial mucin

C:Keywords: alternative splicing; duplication; glycoprotein; glycoprotein; phosphoprotein; polymorphis
F:1-1344/Product: mucin 1 precursor, splice form A #status predicted <PREA>

F:1-62/Region: mucin 1 amino-terminal non-repetitive

F:1-23/Domain: signal sequence #link PREA #status predicted <SIGA>

F:1-19,29-32/Domain: signal sequence #link PREB #status predicted <SIGB>

F:1-19,29-1344/Product: mucin 1 precursor, splice form B #status predicted <PREB>

F:1-19,29-212,1033-1344/Product: mucin 1 precursor, splice form A #status predicted <PREA>

F:138-1017/Region: 20-residue repeats (GSTAPPAGVTSAPDTRPAP)

F:1143-1344/Region: mucin 1 carboxyl-terminal non-repetitive

F:1245-1272/Domain: transmembrane #status predicted <TRN>

F:1046,1064,1118,1144,1222/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:1213/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 100.0%; Score 109; DB 1; Length 1344;
Best Local Similarity 100.0%; Pred. No. 7.2e-06;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HGVTSAPDTRPAGSTAPPA 20
|||||

Db 145 HGVTSAPDTRPAGSTAPPA 164

RESULT 3

A:Residues: 1-19,29-1109,'S',1111-1339,'A',1341-1344 <LAN>
A:Cross-references: UNIPARC:UPI0000174567; GB:J05582; NID:g189598; PIDN:AAA60019.1; PID:
A:Note: GenBank entry HUMPCAB contains four fewer copies of the tandemly repeated sequ
R:Wreschner, D.H.; Hareuveni, M.; Tsarfaty, I.; Smorodinsky, N.; Horev, J.; Zaretzky, J.
Eur. J. Biochem. 189, 463-473, 1990

A:Title: Human epithelial tumor antigen cDNA sequences. Differential splicing may genera
A:Reference number: S10571; MUID:90276413; PMID:2351132

A:Accession: S10572

A:Molecule type: mRNA

A:Residues: 1-19,29-155,'P',157-175,'P',177-182,'A',184-212,1033-1037,'A',1039-1344 <WRE>
A:Cross-references: UNIPARC:UPI0000174569; EMBL:X52229; NID:g37053

R:Wreschner, D.H.
submitted to the EMBL Data Library, March 1990

A:Reference number: S40293

A:Accession: S40293

A:Molecule type: mRNA

A:Residues: 1-19,29-155,'P',157-175,'P',177-182,'A',184-212,1033-1037,'A',1039-1344 <WRE>
A:Cross-references: UNIPARC:UPI0000160A6; EMBL:X52229; NID:g37053; PIDN:CAA36478.1; PID:
R:Abe, M.; Siddiqui, J.; Kufe, D.
Biochem. Biophys. Res. Commun. 165, 644-649, 1989

A:Title: Sequence analysis of the 5' region of the human DF3 breast carcinoma-associated
A:Reference number: A36735; MUID:90088473; PMID:2597151

A:Accession: A36735

A:Molecule type: mRNA

Query Match 85.3%; Score 93; DB 2; Length 256;
Best Local Similarity 85.0%; Pred. No. 0.00015;
Matches 17; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 HGVTSAPDTRPAGSTAPPA 20

A;Accession: F87665
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-267 <STO>
A;Cross-references: UNIPROT:Q9A346; UNIPARC:UPI00000C7A4E; GB:AE005673; NID:g13425064;
C;Genetics:
A;Gene: CC3360

Query Match 51.4%; Score 56; DB 2; Length 267;
Best Local Similarity 64.7%; Pred. No. 5.9;
Matches 11; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 4 TSAPDTRPAPGSTAPPA 20
Db 242 TSVPTDTRAKRSHSAPPA 258
||| ||||| ||| :|||

RESULT 7
AII347
hypothetical protein lmo2185 [imported] - Listeria monocytogenes (strain EGD-e)
C;Species: Listeria monocytogenes
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C;Accession: AII347
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecke
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; M
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AII347
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-569 <GLA>
A;Cross-references: UNIPROT:Q9GW9; UNIPARC:UPI000005571D; GB:NC_003210; PIDN:CAD00263.
A;Experimental source: strain EGD-e
C;Genetics:
A;Gene: lmo2185

Query Match 51.4%; Score 56; DB 2; Length 569;
Best Local Similarity 55.6%; Pred. No. 12;
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 2 GVTSPDTRPAPGSTAPP 19
Db 323 GETNPPVTKPDGTTNPP 340
||| :||| :|||

RESULT 8
T28155
variant-specific surface protein 1 - malaria parasite (Plasmodium falciparum) (fragment
N;Alternate names: erythrocyte membrane binding protein 1 (EMP1)
C;Species: Plasmodium falciparum
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T28155
R;Rowe, J.A.; Moulds, J.M.; Newbold, C.I.; Miller, L.H.
Nature 388, 292-295, 1997
A;Title: Plasmodium falciparum rosetting is mediated by PfEMP1 and requires complement
A;Reference number: Z20477; MUID:97373957; PMID:9230440
A;Accession: T28155
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2706 <ROW>
A;Cross-references: UNIPROT:O15870; UNIPARC:UPI000007886F; EMBL:Y13402; PIDN:CAA73831.1
A;Experimental source: strain IT 4/25/5
C;Genetics:
A;Introns: 2493/3
A;Note: R29R+var1

Query Match 51.4%; Score 56; DB 2; Length 2706;
Best Local Similarity 66.7%; Pred. No. 52;
Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

R:Namba, T.; Oida, H.; Sugimoto, Y.; Kakizuka, A.; Negishi, M.; Ichikawa, A.; Narumiya, J. Biol. Chem. 269, 9986-9992, 1994
A:Title: cDNA cloning of a mouse prostacyclin receptor. Multiple signaling pathways and
A:Reference number: A54416; MUID:94193694; PMID:7511597
A:Accession: A54416
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-417 <NAN>
A:Cross-references: UNIPARC:UPI000016CFCE; GB:D26157; NID:G493687; PIDN:BAA05144.1; PID:
C:Superfamily: prostaglandin E receptor EPI

Query Match 48.6%; Score 53; DB 2; Length 417;
Best Local Similarity 55.0%; Pred. No. 21;
Matches 11; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 1 HGVTAPDTRPAPGSTAPP 20
DB 345 HGDQLPLSRPAGSRDPPA 364

RESULT 12
JC4364
gelatinase B (EC 3.4.24.35) precursor - rat
N:Alternate names: 92-kDa type IV collagenase; matrix metalloproteinase 9 (MMP9)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 10-Jan-1996 #sequence_revision 08-Feb-1996 #text_change 09-Jul-2004
C:Accession: JC4364
R:Okada, A.; Santavica, M.; Basset, P.
Gene 164, 317-321, 1995
A:Title: The cDNA cloning and expression of the gene encoding rat gelatinase B.
A:Reference number: JC4364; MUID:96069602; PMID:7590350
A:Accession: JC4364
A:Molecule type: mRNA
A:Residues: 1-708 <OKA>
A:Cross-references: UNIPROT:P50282; UNIPARC:UPI00001679CE; GB:U24441; NID:g1173505; PIDN:
A:Experimental source: skin wounds
C:Comment: This enzyme belongs to the matrix metalloproteinase family, proteolytic enzym
ogression.
C:Genetics:
A:Gene: gelB
C:Superfamily: gelatinase A; fibronectin type II repeat homology; hemopexin repeat homol
C:Keywords: extracellular matrix; fibroblast; glycoprotein; hydrolase; metalloproteinase
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-708/Product: progelatinase B #status predicted <PRO>
F:25-107/Domain: activation peptide #status predicted <ACT>
F:108-708/Product: gelatinase B #status predicted <MAT>
F:214-389/Region: collagen binding #status predicted
F:231-272/Domain: fibronectin type II repeat homology <2F1>
F:289-330/Domain: fibronectin type II repeat homology <2F2>
F:348-389/Domain: fibronectin type II repeat homology <2F3>
F:514-707/Domain: hemopexin repeat homology <PXN>
F:39,121/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:100,402,406,412/Binding site: zinc, catalytic (Cys, His, His) (inhibited) #status
F:402,406,412/Binding site: zinc, catalytic (His) (active) #status predicted
F:403/Active site: Glu #status predicted
F:519-707/Disulfide bonds: #status predicted

Query Match 48.6%; Score 53; DB 2; Length 708;
Best Local Similarity 55.8%; Pred. No. 35;
Matches 10; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 2 GVTAPDTRPAPGSTAPP 19
DB 483 GPTVAPTGPSPGPTGPP 500

RESULT 13
S62907
gelatinase B (EC 3.4.24.35) precursor - rat
N:Alternate names: collagenase type IV
C:Species: Rattus norvegicus (Norway rat)
C:Date: 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 09-Jul-2004
C:Accession: S62907; S72371

R:Xia, Y.; Garcia, G.; Chen, S.; Wilson, C.B.; Peng, L.
FBES Lett. 382, 285-288, 1996
A:Title: Cloning of rat 92-kDa type IV collagenase and expression of an active recombin
A:Reference number: S62907; MUID:96184505; PMID:8605986
A:Accession: S62907
A>Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-708 <XIA>
A:Cross-references: UNIPROT:P50282; UNIPARC:UPI0000175D91; EMBL:U36476
R:Feng, L.

submitted to the EMBL Data Library, September 1995

A:Reference number: S72371

A:Accession: S72371

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-127,'S',129-708 <FEN>

A:Cross-references: UNIPARC:UPI000012F249; EMBL:U36476; NID:g1022783; PIDN:AAB01721.1; i
C:Superfamily: gelatinase A; fibronectin type II repeat homology; hemopexin repeat homol
C:Keywords: extracellular matrix; fibroblast; glycoprotein; hydrolase; metalloproteinase;

F:1-19/Domain: signal sequence #status predicted <SIG>

F:20-107/Domain: propeptide #status predicted <PRO>

F:68-217,393-445/Domain: matrix metalloproteinase homology #status atypical <MMP>
F:108-708/Product: gelatinase A #status predicted <MAT>
F:214-389/Region: collagen binding #status predicted

F:231-272/Domain: fibronectin type II repeat homology <2F1>

F:289-330/Domain: fibronectin type II repeat homology <2F2>

F:348-389/Domain: fibronectin type II repeat homology <2F3>

F:514-707/Domain: hemopexin repeat homology <PXN>

F:100,402,406,412/Binding site: zinc, catalytic (Cys, His, His) (inhibited) #status

F:402,406,412/Binding site: zinc, catalytic (His) (active) #status predicted

F:403/Active site: Glu #status predicted

F:519-707/Disulfide bonds: #status predicted

Query Match 48.6%; Score 53; DB 2; Length 708;

Best Local Similarity 55.6%; Pred. No. 35;

Matches 10; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 2 GVTAPDTRPAPGSTAPP 19

DB 483 GPTVAPTGPSPGPTGPP 500

RESULT 14

F72570

hypothetical protein APE1847 - Aeropyrum pernix (strain K1)

C:Species: Aeropyrum pernix

C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jul-2004

C:Accession: F72570

R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takai
awa, H.; Takamiya, M.; Maeda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.;

DNA Res. 6, 83-101, 1999

A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropy

A:Reference number: A72450; MUID:99310339; PMID:10382966

A:Accession: F72570

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-115 <KAW>

A:Cross-references: UNIPROT:O9YAU7; UNIPARC:UPI000005E0BF; DDBJ:AP0000062; NID:g5105244;
A:Experimental source: strain K1

C:Genetics:

A:Gene: APE1847

C:Superfamily: Aeropyrum pernix hypothetical protein APE1847

Query Match 47.7%; Score 52; DB 2; Length 115;

Best Local Similarity 58.8%; Pred. No. 8 4;

Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 4 TSAPDTRPAPGSTAPP 20

DB 83 TSTFTTRPGSGTGKPS 99

Tue Apr 18 08:15:26 2006

RESULT 15
S73046
penicillin binding protein pbpC - Mycobacterium leprae
N:Alternate names: L308.f2.77 protein
C:Species: Mycobacterium leprae
C>Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
C:Accession: S73046
R:Smith, D.R.; Robison, K.
submitted to the EMBL Data Library, November 1993
A:Description: Mycobacterium leprae cosmid L308.
A:Reference number: S72590
A:Accession: S73046
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-182 <SMI>
A:Cross-references: UNIPROT:Q49921; UNIPARC:UPI000000BBD98; EMBL:U00022; NID:g467164; PID
C:Genetics:
A:Gene: pbpC

Query Match 47.7%; Score 52; DB 2; Length 182;
Best Local Similarity 47.4%; Pred. No. 13;
Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 2 GVTAPDTRPAPGSTAPPA 20
|:|||||:|
Db 88 GVTAPDTPVPVGDLSAEA 106

Search completed: April 14, 2006, 10:30:26
Job time : 24.4146 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2006, 10:14:04 ; Search time 151.707 Seconds
(without alignments)
93.012 Million cell updates/sec

Title: US-09-606-910E-2

Perfect score: 109
Sequence: 1 HGVTSAPDTRPAPGSTAPPA 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Uniprot 05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	109	100.0	475	2 Q4VX32_HUMAN	Q4VX32 homo sapien
2	109	100.0	1255	1 MUC1_HUMAN	P15941 h mucin-1 p
3	93	85.3	256	2 Q7M4M7_HUMAN	Q7M4M7 homo sapien
4	91	83.5	475	1 MUC1_HYLLA	Q29435 hyllobates l
5	85	78.0	553	2 Q9MZL1_MACMU	Q9MZL1 macaca mula
6	78	71.6	20	2 Q9UM18_HUMAN	Q9UM18 homo sapien
7	63	57.8	745	2 Q89X06_BRAJA	Q89X06 bradyrhizob
8	62	56.9	564	2 Q7F233_ORVSA	Q7F233 oryza sativ
9	62	56.9	1334	2 Q9RKR9_STRCO	Q9RKR9 streptomyc
10	61	56.0	516	2 Q9XVH5_STRPU	Q9XVH5 strongyloce
11	60	55.0	162	2 Q8MJW2_EQUHM	Q8MJW2 equus hemio
12	60	55.0	162	2 Q8MJW4_EQUAS	Q8MJW4 equus asinu
13	60	55.0	168	2 Q8MJV8_EQUZE	Q8MJV8 equus zebra
14	60	55.0	168	2 Q8MJV9_EQUGR	Q8MJV9 equus grevy
15	60	55.0	705	2 Q848C3_STRHY	Q848C3 streptomyc
16	59	54.1	360	2 Q4UW97_XANCP	Q4UW97 xanthomonas
17	59	54.1	360	2 Q8P7V2_XANCP	Q8P7V2 xanthomonas
18	59	54.1	889	2 Q9F2N5_STRCO	Q9F2N5 streptomyc
19	58	53.2	596	2 Q7S5L7_NEUCR	Q7S5L7 neurospora
20	57	52.3	428	2 Q22369_CAEEL	Q22369 caenorhabdi
21	56	51.4	267	2 Q9A346_CAUCR	Q9A346 caulobacter
22	56	51.4	569	2 Q9KGV9_LISMO	Q9KGV9 listeria mo
23	56	51.4	909	2 Q6MP91_BDEBA	Q6MP91 bdellovibri
24	56	51.4	2706	2 Q15870_PLAFA	Q15870 plasmodium
25	55	50.5	329	2 Q9SM15_MAIZE	Q9SM15 zea mays (m
26	55	50.5	400	2 Q73VH2_MYCPA	Q73VH2 mycobacteri
27	55	50.5	429	2 Q61PB3_CAEER	Q61PB3 caenorhabdi
28	55	50.5	569	2 Q71X10_LISMF	Q71X10 listeria mo
29	55	50.5	1366	2 Q4RSK8_TETNG	Q4RSK8 tetraodon n
30	55	50.5	2331	2 Q59EG0_HUMAN	Q59EG0 homo sapien
31	55	50.5	4391	1 PGBM_HUMAN	P98160 homo sapien

32	55	50.5	4391	2 Q5VU27_HUMAN	Q5VU27 homo sapien
33	54	49.5	105	2 Q5JWB5_HUMAN	Q5JWB5 homo sapien
34	54	49.5	299	2 Q4NR67_9DELT	Q4NR67 anaeromyxob
35	54	49.5	363	2 Q6P223_XENTR	Q6P223 xenopus tro
36	54	49.5	484	1 SUFU_HUMAN	Q9UNX1 homo sapien
37	54	49.5	508	2 Q8CD55_MOUSE	Q8CD55 mus musculu
38	54	49.5	526	2 Q4R7R4_MACFA	Q4R7R4 macaca fasc
39	54	49.5	529	1 FGRL1_EAT	Q7TQM3 rattus norv
40	54	49.5	529	2 Q4V8P8_RAT	Q4V8P8 rattus norv
41	54	49.5	533	2 Q7TQE2_MOUSE	Q7TQE2 mus musculu
42	54	49.5	564	1 ZYX_MOUSE	Q62523 mus musculu
43	54	49.5	5141	2 Q700K0_RAT	Q700K0 rattus norv
44	53	48.6	168	2 Q8MJW7_HORSE	Q8MJW7 equus cabal
45	53	48.6	212	2 Q8CC99_MOUSE	Q8CC99 mus musculu

ALIGNMENTS

RESULT 1

Q4VX32_HUMAN
ID Q4VX32_HUMAN PRELIMINARY; PRT; 475 AA.

AC Q4VX32;

DT 13-SEP-2005 (Tremblrel. 31, Created)

DT 13-SEP-2005 (Tremblrel. 31, Last sequence update)

DT 13-SEP-2005 (Tremblrel. 31, Last annotation update)

DE Mucin 1, transmembrane.

GN Name=MUC1; ORFNames=RP11-263K19.2-001;

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae;

OC Homo.

OX NCBI_TaxID=9606;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RA Glihero R;

RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL713999; CAI95078.1; -; Genomic DNA.

KW Transmembrane.

SQ SEQUENCE 475 AA; 49224 MW; CB64D56690818614 CRC64;

Query Match 100.0%; Score 109; DB 2; Length 475;

Best Local Similarity 100.0%; Pred. No. 2.1e-05;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HGVTSAPDTRPAPGSTAPPA 20

|||||

Db 136 HGVTSAPDTRPAPGSTAPPA 155

RESULT 2

MUC1_HUMAN
ID MUC1_HUMAN STANDARD; PRT; 1255 AA.

AC P15941; P13931; P15942; P17626; Q14128; Q14876; Q16437; Q16442;

AC Q16615; Q9BAA4; Q9UE75; Q9UE76; Q9UQL1; Q9Y4J2;

DT 01-JAN-1990 (Rel. 13, Created)

DT 01-APR-1990 (Rel. 14, Last sequence update)

DT 13-SEP-2005 (Rel. 48, Last annotation update)

DE Mucin-1 precursor (MUC-1) (Polymorphic epithelial mucin) (PEMT)

DE (Episialin) (Tumor-associated mucin) (Carcinoma-associated mucin)

DE (tumor-associated epithelial membrane antigen) (EMA) (H23AG) (Peanut-

DE DF3) (CD227 antigen)

GN Name=MUC1;

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae;

OC Homo.

OX NCBI_TaxID=9606;

RN [1]

RP NUCLEOTIDE SEQUENCE (ISOFORM 1).

TISSUE=Pancreas;

MEDLINE=90368716; PubMed=2394722;

Tue Apr 18 08:15:27 2006

us-09-606-910e-2.rup

RA Lan M.S., Batra S.K., Qi W.-N., Metzgar R.S., Hollingsworth M.A.;
 RT "Cloning and sequencing of a human pancreatic tumor mucin cDNA.";
 RL J. Biol. Chem. 265:15294-15299(1990).
 [2]
 RN
 RP NUCLEOTIDE SEQUENCE (ISOFORMS 1 AND 2).
 RX MEDLINE=9020794; PubMed=2318825;
 RA Lichtenberg M.J.L., Vos H.L., Gennissen A.M.C., Hilkens J.;
 RT "Epistatin, a carcinoma-associated mucin, is generated by a
 polymorphic gene encoding splice variants with alternative amino
 termini.";
 RL J. Biol. Chem. 265:5573-5578(1990).
 [3]
 RN
 RP NUCLEOTIDE SEQUENCE (ISOFORM 1).
 RX TISSUE=Mammary carcinoma;
 RC MEDLINE=90368715; PubMed=1697589;
 RA Gendler S.J., Lancaster C.A., Taylor-Papadimitriou J., Duhig T.,
 Peat N., Burchell J., Pemberton L., Lalani E.-N., Wilson D.;
 RT "Molecular cloning and expression of human tumor-associated
 polymorphic epithelial mucin.";
 RL J. Biol. Chem. 265:15286-15293(1990).
 [4]
 RN
 RP NUCLEOTIDE SEQUENCE (ISOFORM 1).
 RX MEDLINE=91097524; PubMed=2268309;
 RA Lancaster C.A., Peat N., Duhig T., Wilson D., Taylor-Papadimitriou J.,
 Gendler S.J.;
 RT "Structure and expression of the human polymorphic epithelial mucin
 gene: an expressed VNTR unit.";
 RL Biochem. Biophys. Res. Commun. 173:1019-1029(1990).
 [5]
 RN
 RP NUCLEOTIDE SEQUENCE (ISOFORM 5).
 RX TISSUE=Mammary carcinoma;
 RC MEDLINE=90276413; PubMed=2351132;
 RA Wreschner D.H., Hareuveni M., Tsarfaty I., Smorodinsky N., Horev J.,
 Zaretsky J., Kotkes P., Weiss M., Lathe R., Dion A., Keydar I.;
 RT "Human epithelial tumor antigen cDNA sequences. Differential splicing
 may generate multiple protein forms.";
 RL Eur. J. Biochem. 189:463-473(1990).
 [6]
 RN
 RP NUCLEOTIDE SEQUENCE.
 RX TISSUE=Mammary carcinoma;
 RC MEDLINE=90276414; PubMed=2212460;
 RA Hareuveni M., Tsarfaty I., Zaretsky J., Kotkes P., Horev J.,
 Zrihan S., Weiss M., Green S., Lathe R., Wreschner D.H.;
 RT "A transcribed gene, containing a variable number of tandem repeats,
 codes for a human epithelial tumor antigen. cDNA cloning, expression
 of the transcribed gene and over-expression in breast cancer tissue.";
 RL Eur. J. Biochem. 189:475-486(1990).
 [7]
 RN
 RP NUCLEOTIDE SEQUENCE (ISOFORM 1).
 RX MEDLINE=91033045; PubMed=1688329; DOI=10.1016/0378-1119(90)90242-J;
 RA Tsarfaty I., Hareuveni M., Horev J., Zaretsky J., Weiss M.,
 Jeltsch J.M., Garnier J.M., Lathe R., Keydar I., Wreschner D.H.;
 RT "Isolation and characterization of an expressed hypersensitive gene
 coding for a breast-cancer-associated antigen.";
 RL Gene 93:313-318(1990).
 [8]
 RN
 RP NUCLEOTIDE SEQUENCE (ISOFORM 7).
 RX MEDLINE=95010060; PubMed=7925397;
 RA Zrihan-Licht S., Vos H.L., Baruch A., Elroy-Stein O., Sagiv D.,
 Keydar I., Hilkens J., Wreschner D.H.;
 RT "Characterization and molecular cloning of a novel MUC1 protein,
 devoid of tandem repeats, expressed in human breast cancer tissue.";
 RL Eur. J. Biochem. 224:787-795(1994).
 [9]
 RN
 RP NUCLEOTIDE SEQUENCE (ISOFORMS 6; 7 AND 8).
 RX MEDLINE=97355747; PubMed=9212228;
 RA Oosterkamp H.M., Scheiner L., Stefanova M.C., Lloyd K.O.,
 Finstad C.L.;
 RT "Comparison of MUC-1 mucin expression in epithelial and non-epithelial
 cancer cell lines and demonstration of a new short variant form (MUC-
 1/2).";
 RL Int. J. Cancer 72:87-94(1997).
 [10]
 RN

RP NUCLEOTIDE SEQUENCE (ISOFORM 7), AND VARIANT MET-1117.
 RA Zhang L.X., Li C.H.;
 RT "Molecular cloning of an isoform of MUC1, MUC1/Y.";
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 [11]
 RN
 RP NUCLEOTIDE SEQUENCE (ISOFORM 9).
 RX TISSUE=Epithelial cancer;
 RC Zhang L.X., Li C.H., Sun L.Y., Yue W.;
 RA "Cloning of a new potential secreted short variant form of MUC1 mucin
 in epithelial cancer cell line.";
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 [12]
 RN
 RP NUCLEOTIDE SEQUENCE [GENOMIC DNA], AND VARIANTS MET-1117 AND ASN-1142.
 RA Livingston R.J., Rieder M.J., Chung M.-W., Ritchie T.K., Olson A.N.,
 Nguyen C.P., Nguyen D.A., Poel C.B., Robertson P.D., Schackwitz W.S.,
 Sherwood J.K., Leithauser B.J., Nickerson D.A.;
 RT "NIHES-SNPs, environmental genome project, NIHES ES15478, Department
 of Genome Sciences, Seattle, WA (URL: http://egp.gs.washington.edu).";
 RL Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.
 [13]
 RN
 RP PARTIAL NUCLEOTIDE SEQUENCE.
 RX MEDLINE=88330762; PubMed=3417635;
 RA Gendler S.J., Taylor-Papadimitriou J., Duhig T., Rothbard J.,
 Burchell J.;
 RT "A highly immunogenic region of a human polymorphic epithelial mucin
 expressed by carcinomas is made up of tandem repeats.";
 RL J. Biol. Chem. 263:12820-12823(1988).
 [14]
 RN
 RP NUCLEOTIDE SEQUENCE OF 1-160 (ISOFORM 2).
 RX MEDLINE=90088473; PubMed=2597151;
 RA Abe M., Siddiqui J., Kufe D.;
 RT "Sequence analysis of the 5' region of the human DF3 breast carcinoma-
 associated antigen gene.";
 RL Biochem. Biophys. Res. Commun. 165:644-649(1989).
 [15]
 RN
 RP NUCLEOTIDE SEQUENCE OF 1-109 (ISOFORM 2).
 RX MEDLINE=96183746; PubMed=8608966;
 RA Weiss M., Baruch A., Keydar I., Wreschner D.H.;
 RT "Preoperative diagnosis of thyroid papillary carcinoma by reverse
 transcriptase polymerase chain reaction of the MUC1 gene.";
 RL Int. J. Cancer 66:55-59(1996).
 [16]
 RN
 RP NUCLEOTIDE SEQUENCE OF 1-89.
 RX TISSUE=Lung;
 RC MEDLINE=96181716; PubMed=8604237;
 RA Yu C.J., Yang P.C., Shew J.Y., Hong T.M., Yang S.C., Lee Y.C.,
 Lee L.N., Luh K.T., Wu C.W.;
 RT "Mucin mRNA expression in lung adenocarcinoma cell lines and
 tissues.";
 RL Oncology 53:118-126(1996).
 [17]
 RN
 RP NUCLEOTIDE SEQUENCE OF 1-46 (ISOFORMS 3 AND 4).
 RX TISSUE=Mammary carcinoma;
 RC Buluwela L., Liu Q., Luqmani Y.A., Gomm J.J., Coombes R.C.;
 RA Submitted (OCT-1992) to the EMBL/GenBank/DBJ databases.
 [18]
 RN
 RP CARBOHYDRATE-LINKAGE SITES WITHIN THE REPEAT.
 RX MEDLINE=97460054; PubMed=9312074; DOI=10.1074/jbc.272.40.24780;
 RA Mueller S., Goletz S., Packer N., Gooley A.A., Lawson A.M.,
 Hanisch F.-G.;
 RT "Localization of O-glycosylation sites on glycopeptide fragments from
 lactation-associated MUC1. All putative sites within the tandem repeat
 are glycosylation targets in vivo.";
 RL J. Biol. Chem. 272:24780-24793(1997).
 [19]
 RN
 RP CARBOHYDRATE-LINKAGE SITES WITHIN THE REPEAT.
 RX MEDLINE=99303572; PubMed=10373415; DOI=10.1074/jbc.274.26.18165;
 RA Mueller S., Alving K., Peter-Katalinic J., Zachara N., Gooley A.A.,
 Hanisch F.-G.;
 RT "High density O-glycosylation on tandem repeat peptide from secretory
 MUC1 of T47D breast cancer cells.";
 RL J. Biol. Chem. 274:18165-18172(1999).
 [20]
 RN

[20]
 RN POLYMORPHISM WITHIN THE REPEAT.
 RX MEDLINE=21359366; PubMed=11350974; DOI=10.1074/jbc.M103187200;
 RA Engelmann K., Balduz S.E., Hantsch F.-G.;
 RT "Identification and topology of variant sequences within individual
 repeat domains of the human epithelial tumor mucin MUC1.";
 RL J. Biol. Chem. 276:27764-27769(2001).
 RN [21]
 RN CHARACTERIZATION OF ISOFORM Y, AND MUTAGENESIS OF ASP-1116.
 RX MEDLINE=99211485; PubMed=10197628;
 RA Baruch A., Hartmann M.-L., Yoeli M., Adereth Y., Greenstein S.,
 RA Stadler Y., Skornik Y., Zaretsky J., Smorodinsky N.I., Keydar I.,
 RA Wreschner D.H.;
 RT "The breast cancer-associated MUC1 gene generates both a receptor and
 its cognate binding protein.";
 RL Cancer Res. 59:1552-1561(1999).
 RN [22]
 RN PARTIAL PROTEIN SEQUENCE, AND CHARACTERIZATION OF CLEAVAGE SITE.
 RX MEDLINE=21240104; PubMed=11341784; DOI=10.1006/bbrc.2001.4775;
 RA Parry S., Silverman H.S., McDermott K., Willis A., Hollingsworth M.A.,
 RA Harris A.;
 RT "Identification of MUC1 proteolytic cleavage sites in vivo.";
 RL Biochem. Biophys. Res. Commun. 283:715-720(2001).
 RN [23]
 RN CHARACTERIZATION.
 RX MEDLINE=21836452; PubMed=11847293; DOI=10.1110/ps.16502;
 RA Wreschner D.H., McGuckin M.A., Williams S.J., Baruch A., Yoeli M.,
 RA Ziv R., Okun L., Zaretsky J., Smorodinsky N., Keydar I., Neophytou P.,
 RA Stacey M., Lin H.-H., Gordon S.;
 RT "Generation of ligand-receptor alliances by 'SEA' module-mediated
 cleavage of membrane-associated mucin proteins.";
 RL Protein Sci. 11:698-706(2002).
 RN [24]
 Query Match 100.0%; Score 109; DB 1; Length 1255;
 Best Local Similarity 100.0%; Pred. No. 5.8e-05;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 HGVTSAPDTRPAPGSTAPPA 20
 DB 136 HGVTSAPDTRPAPGSTAPPA 155
 RESULT 3
 Q7M4M7 HUMAN PRELIMINARY; PRT; 256 AA.
 AC Q7M4M7;
 DT 01-MAR-2004 (TrEMBLrel. 26, Created)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Tumor-associated antigen DF3.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
 OC Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RN NUCLEOTIDE SEQUENCE.
 RX MEDLINE=90058554; PubMed=2582438;
 RA Merlo G.R., Siddiqui J., Cropp C.S., Liscia D.S., Lidereau R.,
 RA Callahan R., Kufe D.W.;
 RT "Frequent alteration of the DF3 tumor-associated antigen gene in
 primary human breast carcinomas.";
 RL Cancer Res. 49:6966-6971(1989).
 DR PIR; A60533;
 DR ENSG00000185499; Homo sapiens.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR GO; GO:0019068; P:virial assembly; IEA.
 DR InterPro; IPR000982; Matrix.
 DR ProDom; PD000741; Matrix; 2.
 SQ SEQUENCE 256 AA; 25766 MW; E0430F0BA7F3B0B4 CRC64;
 Query Match 85.3%; Score 93; DB 2; Length 256;
 Best Local Similarity 85.0%; Pred. No. 0.0011;
 Matches 17; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 HGVTSAPDTRPAPGSTAPPA 20
 DB 19 HGVTSAPDTRPAPGSTAPPA 38
 RESULT 4
 MUC1 HYLLA
 ID MUC1 HYLLA STANDARD; PRT; 475 AA.
 AC Q29435;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 13-SEP-2005 (Rel. 48, Last annotation update)
 DE Mucin-1 precursor (MUC-1).
 GN Name=MUC1;
 OS Hylobates lar (Common gibbon).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 OC Hylobatidae; Hylobates.
 OX NCBI_TaxID=9580;
 RN [1]
 RN NUCLEOTIDE SEQUENCE.
 RX MEDLINE=96351712; PubMed=8747930; DOI=10.1007/BF00292441;
 RA Spicer A.P., Duhig T., Chilton B.S., Gendler S.J.;
 RT "Analysis of mammalian MUC1 genes reveals potential functionally
 important domains.";
 RL Mamm. Genome 6:885-888(1995).
 CC -!- FUNCTION: Direct or indirect interaction with actin cytoskeleton
 (by similarity).
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- PTM: Highly O-glycosylated and probably also N-glycosylated.
 CC -!- SIMILARITY: Contains 1 SEA domain.
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use as long as its content is in no way modified and this statement is not
 removed.
 CC EMBL; L41589; AAA69965.1; -; Genomic DNA.
 CC EMBL; L41625; AAA69918.1; -; Genomic DNA.
 CC EMBL; L41624; AAA69918.1; JOINED; Genomic DNA.
 CC InterPro; IPR000082; SEA.
 CC Pfam; PF01390; SEA; 1.
 CC SMART; SM00200; SEA; 1.
 CC PROSITE; PS50024; SEA; 1.
 KW Actin-binding; Cytoskeleton; Glycoprotein; Repeat; Signal;
 KW Transmembrane.
 FT SIGNAL 1 23 Potential.
 FT CHAIN 24 475 Mucin-1.
 FT TOPO_DOM 24 380 Extracellular (Potential).
 FT TRANSMEM 381 401 Potential.
 FT TOPO_DOM 402 475 Cytoplasmic (Potential).
 FT REPEAT 102 121 1.
 FT REPEAT 122 141 2.
 FT REPEAT 142 161 3.
 FT REPEAT 162 181 4.
 FT DOMAIN 254 371 SEA.
 SQ SEQUENCE 475 AA; 49372 MW; D7A699D6D58C6622 CRC64;
 Query Match 83.5%; Score 91; DB 1; Length 475;
 Best Local Similarity 89.5%; Pred. No. 0.0037;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 HGVTSAPDTRPAPGSTAPPA 19
 DB 136 HGVTSAPDTRPAPGSTAPPA 154
 RESULT 5
 Q9MZL1 MACMU


```

DR GO: GO:0007186; P:G-protein coupled receptor protein signalin. . .; IEA.
DR InterPro; IPR002185; Dopad4 receptor.
DR PRINTS; PR00569; DOPAMINED4R.
KW Receptor.
FT NON_TER 1
FT NON_TER 162
SQ SEQUENCE 162 AA; 16608 MW; 0C742D20FC4A5C37 CRC64;

Query Match 55.0%; Score 60; DB 2; Length 162;
Best Local Similarity 64.7%; Pred. No. 8.5;
Matches 11; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 3 VTSAPDTRPAPGSTAPP 19
DB 87 VTPAPDATPPPPDATAPP 103

RESULT 12
QBMJV4_EQUAS
ID QBMJV4_EQUAS PRELIMINARY; PRT; 162 AA.
AC QBMJV4;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE D4 dopamine receptor (Fragment).
GN Name=DRD4;
OS Equus asinus (Donkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9793;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Hasegawa T., Sato F., Ishida N.;
RT "Determination and variability of nucleotide sequences for D4 dopamine
RT receptor genes (DRD4) in genus Equus.";
RL J. Equine Sci. 13:57-62(2002).
DR EMBL; AB080629; BAC10654.1; -; Genomic DNA.
DR GO: GO:0016021; C:integral to membrane; IEA.
DR GO: GO:0004952; F:dopamine receptor activity; IEA.
DR GO: GO:0004872; F:receptor activity; IEA.
DR GO: GO:0007186; P:G-protein coupled receptor protein signalin. . .; IEA.
DR InterPro; IPR002185; Dopad4 receptor.
DR PRINTS; PR00569; DOPAMINED4R.
KW Receptor.
FT NON_TER 1
FT NON_TER 162
SQ SEQUENCE 162 AA; 16677 MW; 56183D31EA5B3A77 CRC64;

Query Match 55.0%; Score 60; DB 2; Length 162;
Best Local Similarity 64.7%; Pred. No. 8.5;
Matches 11; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 3 VTSAPDTRPAPGSTAPP 19
DB 87 VTPAPDATPPPPDATAPP 103

RESULT 13
QBMJV8_EQUZE
ID QBMJV8_EQUZE PRELIMINARY; PRT; 168 AA.
AC QBMJV8;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE D4 dopamine receptor (Fragment).
GN Name=DRD4;
OS Equus zebra (Mountain zebra).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9791;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Hasegawa T., Sato F., Ishida N.;
RT "Determination and variability of nucleotide sequences for D4 dopamine
RT receptor genes (DRD4) in genus Equus.";
RL J. Equine Sci. 13:57-62(2002).
DR EMBL; AB080629; BAC10654.1; -; Genomic DNA.
DR GO: GO:0016021; C:integral to membrane; IEA.
DR GO: GO:0004952; F:dopamine receptor activity; IEA.
DR GO: GO:0004872; F:receptor activity; IEA.
DR GO: GO:0007186; P:G-protein coupled receptor protein signalin. . .; IEA.
DR InterPro; IPR002185; Dopad4 receptor.
DR PRINTS; PR00569; DOPAMINED4R.
KW Receptor.
FT NON_TER 1
FT NON_TER 162
SQ SEQUENCE 162 AA; 16677 MW; 56183D31EA5B3A77 CRC64;

Query Match 55.0%; Score 60; DB 2; Length 162;
Best Local Similarity 64.7%; Pred. No. 8.5;
Matches 11; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 3 VTSAPDTRPAPGSTAPP 19
DB 87 VTPAPDATPPPPDATAPP 103

RESULT 14
QBMJV9_EQUGR
ID QBMJV9_EQUGR PRELIMINARY; PRT; 168 AA.
AC QBMJV9;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE D4 dopamine receptor (Fragment).
GN Name=DRD4;
OS Equus grevyi (Grevy's zebra).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9792;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Hasegawa T., Sato F., Ishida N.;
RT "Determination and variability of nucleotide sequences for D4 dopamine
RT receptor genes (DRD4) in genus Equus.";
RL J. Equine Sci. 13:57-62(2002).
DR EMBL; AB080634; BAC10659.1; -; Genomic DNA.
DR GO: GO:0016021; C:integral to membrane; IEA.
DR GO: GO:0004952; F:dopamine receptor activity; IEA.
DR GO: GO:0004872; F:receptor activity; IEA.
DR GO: GO:0007186; P:G-protein coupled receptor protein signalin. . .; IEA.
DR InterPro; IPR002185; Dopad4 receptor.
DR PRINTS; PR00569; DOPAMINED4R.
KW Receptor.
FT NON_TER 1
FT NON_TER 168
SQ SEQUENCE 168 AA; 17258 MW; FB12046546D0B185 CRC64;

Query Match 55.0%; Score 60; DB 2; Length 168;
Best Local Similarity 64.7%; Pred. No. 8.8;
Matches 11; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 3 VTSAPDTRPAPGSTAPP 19
DB 87 VTPAPDATPPPPDATAPP 103

RESULT 15
Q848C3_STRHY
ID Q848C3_STRHY PRELIMINARY; PRT; 705 AA.
AC Q848C3;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Type I PKs (Fragment).
GN Name=gdnB;

```


OS Streptomyces hygroscopicus.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1912;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RA Wang Y., Gao Q.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.

DR EMBL; AF521895; AAO61201.1; -; Genomic DNA.

DR GO; GO:0048037; F-cofactor binding; IEA.

DR InterPro; IPR009081; ACP_like.

DR InterPro; IPR002198; ADH_short.

DR InterPro; IPR000183; Decarboxylase2.

DR InterPro; IPR006163; Phosphateth_bind.

DR InterPro; IPR006162; Pantane S.

DR Pfam; PF00106; adh_short; 1.

DR Pfam; PF00550; PP-Binding; 1.

DR PROSITE; PS00075; ACP DOMAIN; 1.

DR PROSITE; PS00879; ODR_DC_2_2; UNKNOWN_1.

DR PROSITE; PS00012; PHOSPHOPANTHETINE; UNKNOWN_1.

FT NON_TER 1

SQ SEQUENCE 705 AA; 72867 MW; CEBD20B0EBC34990 CRC64;

Query Match 55.0%; Score 60; DB 2; Length 705;

Best Local Similarity 52.6%; Pred. No. 39;

Matches 10; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 GVTAPDTPAPGSTAPPA 20

DB 677 GATPSFGATPSFGATLPPA 695

Search completed: April 14, 2006, 10:29:33

Job time : 151.707 secs

This Page Blank (uspto)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2006, 10:29:56 ; Search time 33.6585 Seconds
(without alignments)
49.126 Million cell updates/sec

Title: US-09-606-910E-2

Perfect score: 109
Sequence: 1 HGVTSAPDTRPAPGSTAPPA 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/prodata/1/iaa/5 COMB.pep.*
- 2: /cgn2_6/prodata/1/iaa/6 COMB.pep.*
- 3: /cgn2_6/prodata/1/iaa/H COMB.pep.*
- 4: /cgn2_6/prodata/1/iaa/PCTUS COMB.pep.*
- 5: /cgn2_6/prodata/1/iaa/RE COMB.pep.*
- 6: /cgn2_6/prodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	109	100.0	25	US-09-217-306B-3	Sequence 3, Appli
2	109	100.0	40	US-08-099-354-1	Sequence 1, Appli
3	109	100.0	40	US-08-288-059-7	Sequence 7, Appli
4	105	96.3	24	US-08-737-896-5	Sequence 5, Appli
5	105	96.3	24	US-09-497-232-23	Sequence 23, Appli
6	105	96.3	24	PCT-US96-09951-5	Sequence 5, Appli
7	105	96.3	25	US-09-497-232-5	Sequence 5, Appli
8	105	96.3	30	US-08-737-896-6	Sequence 6, Appli
9	105	96.3	30	PCT-US96-09951-6	Sequence 6, Appli
10	101	92.7	20	US-09-339-944-6	Sequence 6, Appli
11	101	92.7	20	US-09-497-232-10	Sequence 10, Appli
12	101	92.7	20	US-09-651-265-6	Sequence 6, Appli
13	101	92.7	20	US-10-296-317-45	Sequence 45, Appli
14	101	92.7	25	US-08-288-059-28	Sequence 28, Appli
15	101	92.7	30	US-08-134-198E-13	Sequence 13, Appli
16	101	92.7	36	US-10-296-317-64	Sequence 64, Appli
17	101	92.7	38	US-10-296-317-56	Sequence 56, Appli
18	101	92.7	134	US-09-646-028-1	Sequence 1, Appli
19	101	92.7	137	US-09-646-028-2	Sequence 2, Appli
20	101	92.7	138	US-09-646-028-3	Sequence 3, Appli
21	101	92.7	156	US-09-646-028-4	Sequence 4, Appli
22	101	92.7	172	US-09-646-028-49	Sequence 49, Appli
23	101	92.7	177	US-09-646-028-54	Sequence 54, Appli
24	98	89.9	20	US-08-833-807-8	Sequence 8, Appli
25	98	89.9	20	US-09-223-043-8	Sequence 8, Appli
26	98	89.9	20	US-09-291-351-1	Sequence 1, Appli
27	98	89.9	20	US-09-043-731-16	Sequence 16, Appli

28	98	89.9	20	US-09-593-870A-20	Sequence 20, Appli
29	98	89.9	20	US-09-834-240-1	Sequence 1, Appli
30	98	89.9	21	US-08-833-807-7	Sequence 7, Appli
31	98	89.9	21	US-09-223-043-7	Sequence 7, Appli
32	98	89.9	21	US-09-043-731-15	Sequence 15, Appli
33	98	89.9	21	US-09-593-870A-19	Sequence 19, Appli
34	98	89.9	22	US-09-593-870A-46	Sequence 46, Appli
35	97	89.0	24	US-09-217-306B-16	Sequence 16, Appli
36	97	89.0	24	US-09-217-306B-17	Sequence 17, Appli
37	95	87.2	20	US-08-328-536-1	Sequence 1, Appli
38	95	87.2	28	US-08-488-161-9	Sequence 9, Appli
39	95	87.2	28	US-09-273-685-9	Sequence 9, Appli
40	95	87.2	28	PCT-US95-11934-9	Sequence 9, Appli
41	92	84.4	20	US-09-051-685A-7	Sequence 7, Appli
42	91	83.5	20	US-09-497-232-11	Sequence 11, Appli
43	91	83.5	1867	US-08-479-537A-5	Sequence 5, Appli
44	91	83.5	1867	US-09-083-116-5	Sequence 5, Appli
45	91	83.5	1867	US-09-134-916A-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-09-217-306B-3
; Sequence 3, Application US/09217306B
; Patent No. 6465220
; GENERAL INFORMATION:
; APPLICANT: Hassan, Helle
; APPLICANT: Clausen, Henrik
; APPLICANT: Bennett, Eric P.
; TITLE OF INVENTION: Glycosylation Using GalNac-T4 Transferase
; FILE REFERENCE: 8850*1
; CURRENT APPLICATION NUMBER: US/09/217,306B
; CURRENT FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 3
; LENGTH: 25
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: REPEAT
; LOCATION: (1)..(25)
; OTHER INFORMATION: MUC-1 tandem repeat
US-09-217-306B-3

Query Match 100.0%; Score 109; DB 2; Length 25;
Best Local Similarity 100.0%; Pred. No. 1.7e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HGVTSAPDTRPAPGSTAPPA 20
Db 6 HGVTSAPDTRPAPGSTAPPA 25

RESULT 2
US-08-099-354-1
; Sequence 1, Application US/08099354
; Patent No. 5744144
; GENERAL INFORMATION:
; APPLICANT: FINN, OLIVERA J.
; APPLICANT: FONTENOT, J. D.
; APPLICANT: MONTELLARO, RONALD C.
; TITLE OF INVENTION: SYNTHETIC MULTIPLE TANDDEM REPEAT MUCIN
; TITLE OF INVENTION: AND MUCIN-LIKE PEPTIDES, AND USES THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA

ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/099,354
FILING DATE: 30-JUL-1993
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: SIRILLA, GEORGE M.
REGISTRATION NUMBER: 18221
REFERENCE/DOCKET NUMBER: 6137/202246
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-861-3536
TELEFAX: 202-822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 40 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-099-354-1

Query Match 100.0%; Score 109; DB 1; Length 40;
Best Local Similarity 100.0%; Pred. No. 2.8e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HGVTSAPDTRPAPGSTAPPA 20
DB 16 HGVTSAPDTRPAPGSTAPPA 35

RESULT 3
US-08-288-059-7
Sequence 7, Application US/08288059
Patent No. 5827666
GENERAL INFORMATION:
APPLICANT: FINN, OLIVERA J.
APPLICANT: FONTENOT, J. D.
APPLICANT: MONTELLARO, RONALD C.
TITLE OF INVENTION: SYNTHETIC MULTIPLE TANDEM REPEAT MUCIN
TITLE OF INVENTION: AND MUCIN-LIKE PEPTIDES, AND USES THEREOF
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DABBY & CUSHMAN, L.L.P.
STREET: 1100 NEW YORK AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/288,059
FILING DATE: 08-AUG-1994
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: CHAPIN, MARLANA K.
REGISTRATION NUMBER: 35,843
REFERENCE/DOCKET NUMBER: 61137/205204
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-861-3711
TELEFAX: 202-822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:

LENGTH: 40 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-288-059-7
Query Match 100.0%; Score 109; DB 1; Length 40;
Best Local Similarity 100.0%; Pred. No. 2.8e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HGVTSAPDTRPAPGSTAPPA 20
DB 16 HGVTSAPDTRPAPGSTAPPA 35

RESULT 4
US-08-737-896-5
Sequence 5, Application US/08737896
Patent No. 6168804
GENERAL INFORMATION:
APPLICANT: Samuel, John
APPLICANT: Kwon, Glen S.
TITLE OF INVENTION: METHOD FOR ELICITING TH1-SPECIFIC
TITLE OF INVENTION: IMMUNE RESPONSE
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/737,896
FILING DATE: 24-SEP-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/480,499
FILING DATE: 07-JUN-1996
APPLICATION NUMBER: PCT/US96/09551
FILING DATE: 07-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Haile, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07254/037001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 24 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
IMMEDIATE SOURCE:
CLONE: SPQ-065 MUC1
US-08-737-896-5

Query Match 96.3%; Score 105; DB 2; Length 24;
Best Local Similarity 100.0%; Pred. No. 5.3e-07;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HGVTSAPDTRPAPGSTAPP 19
DB 6 HGVTSAPDTRPAPGSTAPP 24

RESULT 5

NAME: Saxe, Bernhard D.
REGISTRATION NUMBER: 28,665
REFERENCE/DOCKET NUMBER: 042881/0114
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 25 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-497-232-5

Query Match 96.3%; Score 105; DB 2; Length 25;
Best Local Similarity 100.0%; Pred. No. 5.5e-07;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HGVTSAPDTRPAPGSTAPP 19
Db 7 HGVTSAPDTRPAPGSTAPP 25

RESULT 8
US-08-737-896-6
; Sequence 6, Application US/08737896
; Patent No. 616804
; GENERAL INFORMATION:
; APPLICANT: Samuel, John
; APPLICANT: Kwon, Glen S.
; TITLE OF INVENTION: METHOD FOR ELICITING TH1-SPECIFIC
; TITLE OF INVENTION: IMMUNE RESPONSE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/737,896
; FILING DATE: 24-SEP-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/480,499
; FILING DATE: 07-JUN-1996
; APPLICATION NUMBER: PCT/US96/09551
; FILING DATE: 07-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07254/037001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; CLONE: SP1-070 MUC1
US-08-737-896-6

Query Match 96.3%; Score 105; DB 2; Length 30;

Best Local Similarity 100.0%; Pred. No. 6.7e-07;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HGVTSAPDTRPAPGSTAPP 19
Db 12 HGVTSAPDTRPAPGSTAPP 30

RESULT 9
PCT-US96-09951-6
; Sequence 6, Application PC/TUS9609951
; GENERAL INFORMATION:
; APPLICANT: The Governors of the University of Alberta
; TITLE OF INVENTION: A METHOD FOR ELICITING A TH1-SPECIFIC
; TITLE OF INVENTION: IMMUNE RESPONSE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: California
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/09951
; FILING DATE: 06-JUN-1996
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Learn, June M.
; REGISTRATION NUMBER: 31,238
; REFERENCE/DOCKET NUMBER: 07254/037WO1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 678-5070
; TELEFAX: (619) 678-5099
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; CLONE: SP1-070 MUC1
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..30
PCT-US96-09951-6

Query Match 96.3%; Score 105; DB 4; Length 30;
Best Local Similarity 100.0%; Pred. No. 6.7e-07;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HGVTSAPDTRPAPGSTAPP 19
Db 12 HGVTSAPDTRPAPGSTAPP 30

RESULT 10
US-09-339-944-6
; Sequence 6, Application US/09339944
; Patent No. 6114129
; GENERAL INFORMATION:
; APPLICANT: AGRAWAL, Babita
; APPLICANT: LONGENECKER, B. Michael
; TITLE OF INVENTION: METHODS OF DETECTING T-CELL ACTIVATION AND TREATING
; TITLE OF INVENTION: DISORDERS ASSOCIATED WITH T-CELL DYSFUNCTION
; FILE REFERENCE: 042881/0129
; CURRENT APPLICATION NUMBER: US/09/339,944

; CURRENT FILING DATE: 1999-06-25
; EARLIER APPLICATION NUMBER: 60/090,916
; EARLIER FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-339-944-6

Query Match 92.7%; Score 101; DB 2; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.4e-06;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GVTSPDTRPAPGSTAPPA 20
Db 1 GVTSPDTRPAPGSTAPPA 19

RESULT 11
US-09-497-232-10
; Sequence 10, Application US/09497232
; Patent No. 6600012
; GENERAL INFORMATION:
; APPLICANT: AGRAWAL, Babita
; KRANTZ, Mark J.
; REDDISH, Mark A.
; LONGENECKER, B. Michael

TITLE OF INVENTION: METHOD FOR GENERATING ACTIVATED T-CELLS
AND ANTIGEN-PULSED ANTIGEN-PRESENTING CELLS

NUMBER OF SEQUENCES: 34

CORRESPONDENCE ADDRESS:

ADDRESSER: FOLEY & LARDNER

STREET: 3000 K Street, N.W.

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/497,232

FILING DATE: 03-Feb-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/074,410

FILING DATE: 08-MAY-1998

APPLICATION NUMBER: US 60/045,949

FILING DATE: 08-MAY-1997

ATTORNEY/AGENT INFORMATION:

NAME: Saxe, Bernhard D.

REGISTRATION NUMBER: 28,665

REFERENCE/DOCKET NUMBER: 042881/0114

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 672-5300

TELEFAX: (202) 672-5399

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 20 amino acids

TYPE: amino acid

STRANDEDNESS: <Unknown>

TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 10:

US-09-497-232-10

Query Match 92.7%; Score 101; DB 2; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.4e-06;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GVTSPDTRPAPGSTAPPA 20
Db 1 GVTSPDTRPAPGSTAPPA 19

RESULT 12
US-09-651-265-6
; Sequence 6, Application US/09651265
; Patent No. 6602660
; GENERAL INFORMATION:
; APPLICANT: AGRAWAL, BABITA
; APPLICANT: LONGENECKER, B. MICHAEL
; TITLE OF INVENTION: METHODS OF DETECTING T-CELL ACTIVATION
; FILE REFERENCE: 042881/0151
; CURRENT APPLICATION NUMBER: US/09/651,265
; CURRENT FILING DATE: 2000-08-30
; PRIOR APPLICATION NUMBER: 09/339,344
; PRIOR FILING DATE: 1999-06-23
; PRIOR APPLICATION NUMBER: 60/090,916
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: peptide
US-09-651-265-6

Query Match 92.7%; Score 101; DB 2; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.4e-06;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GVTSPDTRPAPGSTAPPA 20
Db 1 GVTSPDTRPAPGSTAPPA 19

RESULT 13
US-10-296-317-45
; Sequence 45, Application US/10296317
; Patent No. 6951647
; GENERAL INFORMATION:
; APPLICANT: Cel-Sci Corp
; APPLICANT: Zimmerman, Daniel S
; APPLICANT: Sarin, Prem S
; TITLE OF INVENTION: T CELL BINDING LIGAND PEPTIDES, PEPTIDE
; FILE REFERENCE: CS-112
; CURRENT APPLICATION NUMBER: US/10/296,317
; CURRENT FILING DATE: 2002-11-22
; PRIOR APPLICATION NUMBER: US 60/206548
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: PCT/US07/16793
; PRIOR FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 45
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Mucl Peptide M1c
US-10-296-317-45

Query Match 92.7%; Score 101; DB 2; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.4e-06;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GVTSPDTRPAPGSTAPPA 20

```
Db          1  GVTSAPDTRPAPGSTAPPA 19

RESULT 14
US-08-288-059-28
; Sequence 28, Application US/08288059
; Patent No. 5827666
; GENERAL INFORMATION:
; APPLICANT: FINN, OLIVERA J.
; APPLICANT: FONTENOT, J. D.
; APPLICANT: MONTEJARO, RONALD C.
; TITLE OF INVENTION: SYNTHETIC MULTIPLE TANDEM REPEAT MUCIN
; TITLE OF INVENTION: AND MUCIN-LIKE PEPTIDES, AND USES THEREOF
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DABRY & CUSHMAN, L.L.P.
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/288,059
; FILING DATE: 08-AUG-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: CHAPIN, MARIANA K.
; REGISTRATION NUMBER: 35,843
; REFERENCE/DOCKET NUMBER: 61137/205204
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-861-3711
; TELEFAX: 202-822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-288-059-28

Query Match          92.7%; Score 101; DB 1; Length 25;
Best Local Similarity 100.0%; Pred. No. 1.8e-06;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy          2  GVTSAPDTRPAPGSTAPPA 20
Db          1  GVTSAPDTRPAPGSTAPPA 19

RESULT 15
US-08-134-198E-13
; Sequence 13, Application US/08134198E
; Patent No. 6190885
; GENERAL INFORMATION:
; APPLICANT: CANCER RESEARCH FUND
; APPLICANT: OF CONTRA COSTA
; APPLICANT: PETERSON, JERRY A.
; APPLICANT: LAROCCA, DAVID J.
; TITLE OF INVENTION: FUSION PROTEIN CONTAINING HMFG
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder & Poplawski
; STREET: 444 South Flower Street, Suite 1900
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
```

```
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0,
; SOFTWARE: Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/134,198E
; FILING DATE: October 8, 1993
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Amzel, Viviana
; REGISTRATION NUMBER: 30,930
; REFERENCE/DOCKET NUMBER: P66 38208 (CRFC-003C)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 622-7700
; TELEFAX: (213) 489-4210
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-134-198E-13

Query Match          92.7%; Score 101; DB 2; Length 30;
Best Local Similarity 100.0%; Pred. No. 2.1e-06;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy          2  GVTSAPDTRPAPGSTAPPA 20
Db          1  GVTSAPDTRPAPGSTAPPA 19

Search completed: April 14, 2006, 10:31:39
Job time : 33.6585 secs
```


GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2006, 10:45:37 ; Search time 113.171 Seconds
(without alignments)
73.841 Million cell updates/sec

Title: US-09-606-910E-2

Perfect score: 109

Sequence: 1 HGVTSAPDTRPAGSTAPPA 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications AA Main:

1: /cgm2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgm2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
3: /cgm2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
4: /cgm2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
5: /cgm2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
6: /cgm2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	109	100.0	20	4	US-10-292-896-1
2	109	100.0	20	5	US-10-705-401-1
3	109	100.0	25	3	US-09-881-339-3
4	109	100.0	25	4	US-10-417-633-3
5	109	100.0	40	3	US-09-996-069-9
6	109	100.0	40	5	US-10-635-211-4
7	109	100.0	43	4	US-10-106-876-14
8	109	100.0	46	3	US-09-996-069-5
9	109	100.0	46	4	US-10-106-876-1
10	109	100.0	100	3	US-09-965-131-6
11	109	100.0	159	5	US-10-635-211-9
12	109	100.0	475	4	US-10-417-312-1
13	109	100.0	475	5	US-10-696-633-59
14	109	100.0	475	6	US-10-756-149-5038
15	109	100.0	475	6	US-11-055-119-67
16	109	100.0	495	6	US-11-055-119-2
17	109	100.0	508	4	US-10-057-136-20
18	109	100.0	515	4	US-10-097-340-212
19	109	100.0	515	4	US-10-171-311-156
20	109	100.0	515	4	US-10-612-090-19
21	109	100.0	515	6	US-11-050-926-212
22	109	100.0	586	5	US-10-635-211-2
23	109	100.0	1255	3	US-09-996-069-10
24	109	100.0	1255	4	US-10-171-311-158
25	109	100.0	1255	4	US-10-177-293-311
26	109	100.0	1255	4	US-10-734-564-120
27	109	100.0	1255	5	US-10-473-484-2

28 109 100.0 1255 5 US-10-997-055-2 Sequence 2, Appli
29 109 100.0 1255 6 US-11-037-713-20 Sequence 20, Appli
30 109 100.0 1255 6 US-11-009-533-2 Sequence 2, Appli
31 105 96.3 24 4 US-10-292-896-2 Sequence 2, Appli
32 105 96.3 24 5 US-10-705-401-2 Sequence 5, Appli
33 105 96.3 26 3 US-09-815-346-5 Sequence 8, Appli
34 105 96.3 26 4 US-10-106-876-8 Sequence 20, Appli
35 105 96.3 26 4 US-10-106-876-20 Sequence 4, Appli
36 105 96.3 27 3 US-09-815-346-4 Sequence 1, Appli
37 105 96.3 27 3 US-09-870-691-1 Sequence 1, Appli
38 105 96.3 27 3 US-09-999-191-1 Sequence 7, Appli
39 105 96.3 27 4 US-10-106-876-7 Sequence 9, Appli
40 105 96.3 27 4 US-10-106-876-9 Sequence 12, Appli
41 105 96.3 27 4 US-10-106-876-12 Sequence 13, Appli
42 105 96.3 28 4 US-10-106-876-13 Sequence 16, Appli
43 101 92.7 20 3 US-09-984-183-16 Sequence 6, Appli
44 101 92.7 20 3 US-09-984-333-6 Sequence 1, Appli
45 101 92.7 20 4 US-10-054-488-1

ALIGNMENTS

RESULT 1

US-10-292-896-1

; Sequence 1, Application US/10292896

; Publication No. US20030186850A1

; GENERAL INFORMATION:

; APPLICANT: HASSAN, Helle

; APPLICANT: REIS, Celso A.

; APPLICANT: BENNETT, Eric P.

; TITLE OF INVENTION: METHODS TO IDENTIFY AGENTS MODULATING FUNCTIONS OF POLYPEPTIDE G.

; TITLE OF INVENTION: TRANSFERASES, PHARMACEUTICAL COMPOSITIONS COMPRISING SUCH AGENT

; TITLE OF INVENTION: SUCH AGENTS FOR PREPARING MEDICAMENTS

; FILE REFERENCE: 4305/1H154-US3

; CURRENT APPLICATION NUMBER: US/10/292,896

; CURRENT FILING DATE: 2003-03-17

; PRIOR APPLICATION NUMBER: US 60/425,204

; PRIOR FILING DATE: 2002-11-08

; PRIOR APPLICATION NUMBER: PCT/DK01/00328

; PRIOR FILING DATE: 2001-05-10

; PRIOR APPLICATION NUMBER: US 60/203,331

; PRIOR FILING DATE: 2000-05-11

; NUMBER OF SEQ ID NOS: 130

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 1

; LENGTH: 20

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-292-896-1

Query Match 100.0%; Score 109; DB 4; Length 20;

Best Local Similarity 100.0%; Pred. No. 1.3e-06;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HGVTSAPDTRPAGSTAPPA 20

Db 1 HGVTSAPDTRPAGSTAPPA 20

RESULT 2

US-10-705-401-1

; Sequence 1, Application US/10705401

; Publication No. US20050026266A1

; GENERAL INFORMATION:

; APPLICANT: Clausen, Henrik

; APPLICANT: Bennett, Eric P.

; TITLE OF INVENTION: METHODS TO IDENTIFY AGENTS MODULATING FUNCTIONS OF POLYPEPTIDE G.

; TITLE OF INVENTION: TRANSFERASES, PHARMACEUTICAL COMPOSITIONS COMPRISING SUCH AGENT

; TITLE OF INVENTION: USE OF SUCH AGENTS FOR PREPARING MEDICAMENTS

; FILE REFERENCE: 04305/100H154-US2

; CURRENT APPLICATION NUMBER: US/10/705,401

; CURRENT FILING DATE: 2003-11-10
; PRIOR APPLICATION NUMBER: US 60/425,204
; PRIOR FILING DATE: 2002-11-08
; PRIOR APPLICATION NUMBER: PCT/DK03/00763
; PRIOR FILING DATE: 2003-11-07
; NUMBER OF SEQ ID NOS: 127
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: synthetic peptide
US-10-705-401-1

Query Match 100.0%; Score 109; DB 5; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.3e-06;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HGVTSA PDTRPAGSTAPPA 20
|||||
Db 1 HGVTSA PDTRPAGSTAPPA 20

RESULT 3

US-09-881-339-3
; Sequence 3, Application US/09881339
; Publication No. US20030138860A1
; GENERAL INFORMATION:
; APPLICANT: Robertson, John F.R.
; APPLICANT: Graves, Catherine
; APPLICANT: Price, Michael R.
; TITLE OF INVENTION: Cancer Detection Methods and Reagents
; FILE REFERENCE: 02332-0030 (49409-264876)
; CURRENT APPLICATION NUMBER: US/09/881,339
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,886
; PRIOR FILING DATE: 2000-06-14
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 25
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic MUC1 peptide TAP2

; NAME/KEY: MISC FEATURE
; LOCATION: (9)..(9)
; OTHER INFORMATION: T is O-glycosylated with N-acetylgalactosamine
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (21)..(21)
; OTHER INFORMATION: T is O-glycosylated with N-acetylgalactosamine
US-09-881-339-3

Query Match 100.0%; Score 109; DB 3; Length 25;
Best Local Similarity 100.0%; Pred. No. 1.6e-06;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HGVTSA PDTRPAGSTAPPA 20
|||||
Db 6 HGVTSA PDTRPAGSTAPPA 25

RESULT 4

US-10-417-633-3
; Sequence 3, Application US/10417633
; Publication No. US20030232399A1
; GENERAL INFORMATION:
; APPLICANT: Robertson, John F.R.
; APPLICANT: Graves, Catherine
; APPLICANT: Price, Michael R.

; TITLE OF INVENTION: Cancer Detection Methods and Reagents
; FILE REFERENCE: 52623-0031 52623-284706
; CURRENT APPLICATION NUMBER: US/10/417,633
; CURRENT FILING DATE: 2003-04-16
; PRIOR APPLICATION NUMBER: US 60/211,886
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 09/881,339
; PRIOR FILING DATE: 2001-06-14
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 25
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic MUC1 peptide TAP2
; NAME/KEY: MISC FEATURE
; LOCATION: (9)..(9)
; OTHER INFORMATION: T is O-glycosylated with N-acetylgalactosamine
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (21)..(21)
; OTHER INFORMATION: T is O-glycosylated with N-acetylgalactosamine
US-10-417-633-3

Query Match 100.0%; Score 109; DB 4; Length 25;
Best Local Similarity 100.0%; Pred. No. 1.6e-06;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HGVTSA PDTRPAGSTAPPA 20
|||||
Db 6 HGVTSA PDTRPAGSTAPPA 25

RESULT 5

US-09-996-069-9
; Sequence 9, Application US/09996069
; Publication No. US20030036199A1
; GENERAL INFORMATION:
; APPLICANT: Bamdad, R. Shoshana
; APPLICANT: Bamdad, Cynthia
; TITLE OF INVENTION: DIAGNOSTIC TUMOR MARKERS, DRUG SCREENING FOR TUMORIGENESIS INHIBITORS
; FILE REFERENCE: M01015/70071
; CURRENT APPLICATION NUMBER: US/09/996,069
; CURRENT FILING DATE: 2001-11-27
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 40
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-996-069-9

Query Match 100.0%; Score 109; DB 3; Length 40;
Best Local Similarity 100.0%; Pred. No. 2.5e-06;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HGVTSA PDTRPAGSTAPPA 20
|||||
Db 15 HGVTSA PDTRPAGSTAPPA 34

RESULT 6

US-10-635-211-4
; Sequence 4, Application US/10635211
; Publication No. US20050031649A1
; GENERAL INFORMATION:
; APPLICANT: Beijing HYD VAX Biotechnology Co. Ltd
; TITLE OF INVENTION: A recombinant fusion protein comprising BCG heat shock protein 6
; TITLE OF INVENTION: and the epitope of MUC1
; FILE REFERENCE: FP03012US

```

: APPLICANT: O'Donnell, Mark A.
: TITLE OF INVENTION: RECOMBINANT BCG VACCINES FOR THE
: TITLE OF INVENTION: PREVENTION AND TREATMENT OF CANCER
: FILE REFERENCE: WII-014CP
: CURRENT APPLICATION NUMBER: US/09/965,131
: CURRENT FILING DATE: 2001-09-26
: PRIOR APPLICATION NUMBER: 60/235,455
: PRIOR FILING DATE: 2000-09-26
: NUMBER OF SEQ ID NOS: 6
: SOFTWARE: FastSEQ for Windows Version 4.0
: SEQ ID NO 6
: LENGTH: 100
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-965-131-6

Query March 100.0%; Score 109; DB 3; Length 100;
Best Local Similarity 100.0%; Pred. No. 6.3e-06;

```

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HGVTSAPDTRPAPGSTAPPA 20
| | | | | | | | | | | | | | | | | | | | | |
Db 20 HGVTSAPDTRPAPGSTAPPA 39

RESULT 11

US-10-635-211-9
; Sequence 9, Application US/10635211
; Publication No. US20050031649A1
; GENERAL INFORMATION:
; APPLICANT: Beijing HYDVAX Biotechnology Co. Ltd
; TITLE OF INVENTION: A recombinant fusion protein comprising BCG heat shock protein 65
; FILE REFERENCE: FP03012US
; CURRENT APPLICATION NUMBER: US/10/635,211
; CURRENT FILING DATE: 2003-08-06
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
; LENGTH: 159
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion Protein
US-10-635-211-9

Query Match 100.0%; Score 109; DB 5; Length 159;
Best Local Similarity 100.0%; Pred. No. 1e-05;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HGVTSAPDTRPAPGSTAPPA 20
| | | | | | | | | | | | | | | | | | | | | |
Db 21 HGVTSAPDTRPAPGSTAPPA 40

RESULT 12

US-10-417-312-1
; Sequence 1, Application US/10417312
; Publication No. US2003023568A1
; GENERAL INFORMATION:
; APPLICANT: Dyax Corp
; TITLE OF INVENTION: Antibodies Specific for Mucin Polypeptide
; FILE REFERENCE: 2403/2002
; CURRENT APPLICATION NUMBER: US/10/417,312
; CURRENT FILING DATE: 2003-04-16
; PRIOR APPLICATION NUMBER: US 60/374,432
; PRIOR FILING DATE: 2002-04-22
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 475
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-417-312-1

Query Match 100.0%; Score 109; DB 4; Length 475;
Best Local Similarity 100.0%; Pred. No. 3e-05;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HGVTSAPDTRPAPGSTAPPA 20
| | | | | | | | | | | | | | | | | | | | | |
Db 136 HGVTSAPDTRPAPGSTAPPA 155

RESULT 13

US-10-696-639-59
; Sequence 59, Application US/10696639
; Publication No. US20050037439A1
; GENERAL INFORMATION:
; APPLICANT: Pharmacia Corporation
; APPLICANT: Bournier, Maureen J.

; TITLE OF INVENTION: DIFFERENTIALLY EXPRESSED GENES INVOLVED IN CANCER, THE
; TITLE OF INVENTION: POLYPEPTIDES ENCODED THEREBY, AND METHODS OF USING THE SAME
; FILE REFERENCE: 01040/1
; CURRENT APPLICATION NUMBER: US/10/696,639
; CURRENT FILING DATE: 2003-10-29
; PRIOR APPLICATION NUMBER: 60/422,176
; PRIOR FILING DATE: 2002-10-29
; NUMBER OF SEQ ID NOS: 3114
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 59
; LENGTH: 475
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-696-639-59

Query Match 100.0%; Score 109; DB 5; Length 475;
Best Local Similarity 100.0%; Pred. No. 3e-05;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HGVTSAPDTRPAPGSTAPPA 20
| | | | | | | | | | | | | | | | | | | | | |
Db 136 HGVTSAPDTRPAPGSTAPPA 155

RESULT 14

US-10-756-149-5038
; Sequence 5038, Application US/10756149
; Publication No. US20050181375A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSIS OF METASTATIC CANCER, COMPOSITIONS AND
; TITLE OF INVENTION: METHODS OF SCREENING FOR MODULATORS OF METASTATIC CANCER
; FILE REFERENCE: file
; CURRENT APPLICATION NUMBER: US/10/756,149
; CURRENT FILING DATE: 2004-01-12
; NUMBER OF SEQ ID NOS: 5818
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5038
; LENGTH: 475
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-756-149-5038

Query Match 100.0%; Score 109; DB 5; Length 475;
Best Local Similarity 100.0%; Pred. No. 3e-05;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HGVTSAPDTRPAPGSTAPPA 20
| | | | | | | | | | | | | | | | | | | | | |
Db 136 HGVTSAPDTRPAPGSTAPPA 155

RESULT 15

US-11-055-119-67
; Sequence 67, Application US/11055119
; Publication No. US20050142640A1
; GENERAL INFORMATION:
; APPLICANT: Taylor-Papadimitriou, Joyce
; APPLICANT: Heukamp, Lukas Carl
; APPLICANT: Offringa, Rienk
; APPLICANT: Melief, Cornelis Johanna Maria
; APPLICANT: Acres, Bruce
; APPLICANT: Thomas, Mireille
; TITLE OF INVENTION: MUC-1 derived peptides
; FILE REFERENCE: 029395-017
; CURRENT APPLICATION NUMBER: US/11/055,119
; CURRENT FILING DATE: 2005-02-11
; PRIOR APPLICATION NUMBER: US/09/658,621
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: US 60/187,215
; PRIOR FILING DATE: 2000-03-03
; PRIOR APPLICATION NUMBER: GB 9921242.5

```
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: EP 99 40 2237.4
; PRIOR FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 67
; LENGTH: 475
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-055-119-67

Query Match      100.0%; Score 109; DB 6; Length 475;
Best Local Similarity 100.0%; Pred. No. 3e-05;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 HGVTSAPDTRPAPGSTAPPA 20
      |||||||
Db      136 HGVTSAPDTRPAPGSTAPPA 155

Search completed: April 14, 2006, 10:50:38
Job time : 114.171 secs
```

This Page Blank (uspto)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2006, 10:46:57 ; Search time 17.0732 Seconds
(without alignments)
49.773 Million cell updates/sec

Title: US-09-606-910E-2

Perfect score: 109
Sequence: 1 HGVTSAPDTRPAGSTAPPA 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 217505 seqs, 42489236 residues

Total number of hits satisfying chosen parameters: 217505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Published Applications AA_New.*
- 1: /SIDSS/ptodata/2/pubppa/US08_NEW_PUB.pep.*
 - 2: /SIDSS/ptodata/2/pubppa/US06_NEW_PUB.pep.*
 - 3: /SIDSS/ptodata/2/pubppa/US07_NEW_PUB.pep.*
 - 4: /SIDSS/ptodata/2/pubppa/PCT_NEW_PUB.pep.*
 - 5: /SIDSS/ptodata/2/pubppa/US09_NEW_PUB.pep.*
 - 6: /SIDSS/ptodata/2/pubppa/US10_NEW_PUB.pep.*
 - 7: /SIDSS/ptodata/2/pubppa/US11_NEW_PUB.pep.*
 - 8: /SIDSS/ptodata/2/pubppa/US60_NEW_PUB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	109	100.0	20	7	US-11-033-365-138
2	109	100.0	20	7	US-11-033-365-139
3	109	100.0	21	7	US-11-088-546-8
4	109	100.0	109	6	US-10-517-696-115
5	109	100.0	316	6	US-10-517-696-126
6	109	100.0	325	6	US-10-517-696-148
7	109	100.0	336	6	US-10-517-696-153
8	109	100.0	350	6	US-10-517-696-125
9	109	100.0	372	6	US-10-517-696-129
10	109	100.0	379	6	US-10-517-696-150
11	109	100.0	396	6	US-10-517-696-147
12	109	100.0	398	6	US-10-517-696-136
13	109	100.0	409	6	US-10-517-696-149
14	109	100.0	420	6	US-10-517-696-141
15	109	100.0	463	6	US-10-517-696-121
16	109	100.0	475	6	US-10-501-033-309
17	109	100.0	475	6	US-10-517-696-117
18	109	100.0	475	6	US-10-515-872-10
19	109	100.0	483	6	US-10-517-696-132
20	109	100.0	484	6	US-10-517-696-120
21	109	100.0	485	6	US-10-517-696-142
22	109	100.0	492	6	US-10-517-696-131
23	109	100.0	515	6	US-10-517-696-114
24	109	100.0	515	7	US-11-233-510-20
25	109	100.0	517	6	US-10-517-696-144

Sequence 113, App
Sequence 122, App
Sequence 128, App
Sequence 112, App
Sequence 487, App
Sequence 398, App
Sequence 111, App
Sequence 140, App
Sequence 44, Appl
Sequence 31, Appl
Sequence 14, Appl
Sequence 116, App
Sequence 130, App
Sequence 123, App
Sequence 4, Appl
Sequence 50, Appl
Sequence 7, Appl
Sequence 33, Appl
Sequence 34, Appl
Sequence 8, Appl

ALIGNMENTS

RESULT 1
US-11-033-365-138
; Sequence 138, Application US/11033365
; Publication No. US20050250678A1
; GENERAL INFORMATION:
; APPLICANT: Neose Technologies Inc.
; APPLICANT: DeFrees, Shawn
; APPLICANT: Zopf, David
; APPLICANT: Wang, ZhiGuang
; APPLICANT: Clausen, Henrik
; TITLE OF INVENTION: O-Linked Glycosylation of peptides
; FILE REFERENCE: 040853-01-5138
; CURRENT APPLICATION NUMBER: US/11/033,365
; CURRENT FILING DATE: 2005-01-10
; PRIOR APPLICATION NUMBER: 60/535,284
; PRIOR FILING DATE: 2004-01-08
; PRIOR APPLICATION NUMBER: 60/544,411
; PRIOR FILING DATE: 2004-02-12
; PRIOR APPLICATION NUMBER: 60/546,631
; PRIOR FILING DATE: 2004-02-20
; PRIOR APPLICATION NUMBER: 60/555,813
; PRIOR FILING DATE: 2004-03-23
; PRIOR APPLICATION NUMBER: 60/570,891
; PRIOR FILING DATE: 2004-05-12
; NUMBER OF SEQ ID NOS: 213
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 138
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-033-365-138

Query Match 100.0%; Score 109; DB 7; Length 20;
Best Local Similarity 100.0%; Pred. No. 1e-08; 0;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HGVTSAPDTRPAGSTAPPA 20
|||||
DB 1 HGVTSAPDTRPAGSTAPPA 20

RESULT 2
US-11-033-365-139
; Sequence 139, Application US/11033365
; Publication No. US20050250678A1
; GENERAL INFORMATION:
; APPLICANT: Neose Technologies Inc.
; APPLICANT: DeFrees, Shawn

Tue Apr 18 08:15:26 2006

```
; APPLICANT: Zopf, David
; APPLICANT: Wang, ZhiGuang
; APPLICANT: Clausen, Henrik
; TITLE OF INVENTION: O-Linked Glycosylation of peptides
; FILE REFERENCE: 040853-01-5138
; CURRENT APPLICATION NUMBER: US/11/033,365
; CURRENT FILING DATE: 2005-01-10
; PRIOR APPLICATION NUMBER: 60/535,284
; PRIOR FILING DATE: 2004-01-08
; PRIOR APPLICATION NUMBER: 60/544,411
; PRIOR FILING DATE: 2004-02-12
; PRIOR APPLICATION NUMBER: 60/546,631
; PRIOR FILING DATE: 2004-02-20
; PRIOR APPLICATION NUMBER: 60/555,813
; PRIOR FILING DATE: 2004-03-23
; PRIOR APPLICATION NUMBER: 60/570,891
; PRIOR FILING DATE: 2004-05-12
; NUMBER OF SEQ ID NOS: 213
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 139
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-033-365-139

Query Match      100.0%; Score 109; DB 7; Length 20;
Best Local Similarity 100.0%; Pred. No. 1e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HGVTSAPDTRPAPGSTAPPA 20
Db 1 HGVTSAPDTRPAPGSTAPPA 20

RESULT 3
US-11-088-546-8
; Sequence 8, Application US/11088546
; Publication No. US20060063736A1
; GENERAL INFORMATION:
; APPLICANT: Carolyn R. Bertozzi
; APPLICANT: Howard C. Hang
; TITLE OF INVENTION: Compositions and methods for inhibiting
; TITLE OF INVENTION: mucin-type O-linked glycosylation
; FILE REFERENCE: BERK-028
; CURRENT APPLICATION NUMBER: US/11/088,546
; CURRENT FILING DATE: 2005-03-23
; PRIOR APPLICATION NUMBER: 60/556,673
; PRIOR FILING DATE: 2004-03-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: ppGalNAcT substrate
US-11-088-546-8

Query Match      100.0%; Score 109; DB 7; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.1e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HGVTSAPDTRPAPGSTAPPA 20
Db 2 HGVTSAPDTRPAPGSTAPPA 21

RESULT 4
US-10-517-696-115
; Sequence 115, Application US/10517696
; Publication No. US20060051759A1
; GENERAL INFORMATION:
; APPLICANT: diaDexus, Inc.
```

```
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto A.
; APPLICANT: Turner, Leah R.
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenchua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and Pr
; FILE REFERENCE: DEX-0432
; CURRENT APPLICATION NUMBER: US/10/517,696
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 60/389,327
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 171
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 115
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-517-696-115

Query Match      100.0%; Score 109; DB 6; Length 109;
Best Local Similarity 100.0%; Pred. No. 5.9e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HGVTSAPDTRPAPGSTAPPA 20
Db 40 HGVTSAPDTRPAPGSTAPPA 59

RESULT 5
US-10-517-696-126
; Sequence 126, Application US/10517696
; Publication No. US20060051759A1
; GENERAL INFORMATION:
; APPLICANT: diaDexus, Inc.
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto A.
; APPLICANT: Turner, Leah R.
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenchua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and Pr
; FILE REFERENCE: DEX-0432
; CURRENT APPLICATION NUMBER: US/10/517,696
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 60/389,327
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 171
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 126
; LENGTH: 316
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-517-696-126

Query Match      100.0%; Score 109; DB 6; Length 316;
Best Local Similarity 100.0%; Pred. No. 1.8e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HGVTSAPDTRPAPGSTAPPA 20
Db 145 HGVTSAPDTRPAPGSTAPPA 164

RESULT 6
US-10-517-696-148
; Sequence 148, Application US/10517696
; Publication No. US20060051759A1
; GENERAL INFORMATION:
; APPLICANT: diaDexus, Inc.
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto A.
; APPLICANT: Turner, Leah R.
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenchua
```



```
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and P
; FILE REFERENCE: DEX-0432
; CURRENT APPLICATION NUMBER: US/10/517,696
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 60/389,327
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 171
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 148
; LENGTH: 325
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-517-696-148

Query Match 100.0%; Score 109; DB 6; Length 325;
Best Local Similarity 100.0%; Pred. No. 1.8e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HGVTSAPDTRPAPGSTAPPA 20
DB 145 HGVTSAPDTRPAPGSTAPPA 164

RESULT 7
US-10-517-696-153
; Sequence 153, Application US/10517696
; Publication No. US20060051759A1
; GENERAL INFORMATION:
; APPLICANT: diaDexus, Inc.
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto A.
; APPLICANT: Turner, Leah R.
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chonghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and P
; FILE REFERENCE: DEX-0432
; CURRENT APPLICATION NUMBER: US/10/517,696
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 60/389,327
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 171
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 153
; LENGTH: 336
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-517-696-153

Query Match 100.0%; Score 109; DB 6; Length 336;
Best Local Similarity 100.0%; Pred. No. 1.9e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HGVTSAPDTRPAPGSTAPPA 20
DB 145 HGVTSAPDTRPAPGSTAPPA 164

RESULT 8
US-10-517-696-125
; Sequence 125, Application US/10517696
; Publication No. US20060051759A1
; GENERAL INFORMATION:
; APPLICANT: diaDexus, Inc.
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto A.
; APPLICANT: Turner, Leah R.
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chonghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and P
; FILE REFERENCE: DEX-0432
; CURRENT APPLICATION NUMBER: US/10/517,696
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 60/389,327
```

```
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 171
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 125
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-517-696-125
```

```
Query Match 100.0%; Score 109; DB 6; Length 350;
Best Local Similarity 100.0%; Pred. No. 2e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 HGVTSAPDTRPAPGSTAPPA 20
DB 145 HGVTSAPDTRPAPGSTAPPA 164
```

```
RESULT 9
US-10-517-696-129
; Sequence 129, Application US/10517696
; Publication No. US20060051759A1
; GENERAL INFORMATION:
; APPLICANT: diaDexus, Inc.
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto A.
; APPLICANT: Turner, Leah R.
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chonghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and P
; FILE REFERENCE: DEX-0432
; CURRENT APPLICATION NUMBER: US/10/517,696
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 60/389,327
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 171
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 129
; LENGTH: 372
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-517-696-129
```

```
Query Match 100.0%; Score 109; DB 6; Length 372;
Best Local Similarity 100.0%; Pred. No. 2.1e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 HGVTSAPDTRPAPGSTAPPA 20
DB 145 HGVTSAPDTRPAPGSTAPPA 164
```

```
RESULT 10
US-10-517-696-150
; Sequence 150, Application US/10517696
; Publication No. US20060051759A1
; GENERAL INFORMATION:
; APPLICANT: diaDexus, Inc.
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto A.
; APPLICANT: Turner, Leah R.
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chonghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and P
; FILE REFERENCE: DEX-0432
; CURRENT APPLICATION NUMBER: US/10/517,696
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 60/389,327
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 171
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 150
; LENGTH: 379
```

```

; TYPE: PRT
; ORGANISM: Homo sapien
US-10-517-696-150

Query Match      100.0%; Score 109; DB 6; Length 379;
Best Local Similarity 100.0%; Pred. No. 2.1e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HGVTSAPDTRPAPGSTAPPA 20
Db 145 HGVTSAPDTRPAPGSTAPPA 164

RESULT 11
US-10-517-696-147
; Sequence 147, Application US/10517696
; Publication No. US20060051759A1
; GENERAL INFORMATION:
; APPLICANT: diadexus, Inc.
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto A.
; APPLICANT: Turner, Leah R.
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and Pr
; FILE REFERENCE: DEX-0432
; CURRENT APPLICATION NUMBER: US/10/517,696
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 60/389,327
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 171
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 147
; LENGTH: 396
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-517-696-147

Query Match      100.0%; Score 109; DB 6; Length 396;
Best Local Similarity 100.0%; Pred. No. 2.2e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HGVTSAPDTRPAPGSTAPPA 20
Db 145 HGVTSAPDTRPAPGSTAPPA 164

RESULT 12
US-10-517-696-136
; Sequence 136, Application US/10517696
; Publication No. US20060051759A1
; GENERAL INFORMATION:
; APPLICANT: diadexus, Inc.
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto A.
; APPLICANT: Turner, Leah R.
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and Pr
; FILE REFERENCE: DEX-0432
; CURRENT APPLICATION NUMBER: US/10/517,696
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 60/389,327
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 171
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 136
; LENGTH: 398
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-517-696-136

Query Match      100.0%; Score 109; DB 6; Length 398;
Best Local Similarity 100.0%; Pred. No. 2.2e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HGVTSAPDTRPAPGSTAPPA 20
Db 145 HGVTSAPDTRPAPGSTAPPA 164

RESULT 13
US-10-517-696-149
; Sequence 149, Application US/10517696
; Publication No. US20060051759A1
; GENERAL INFORMATION:
; APPLICANT: diadexus, Inc.
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto A.
; APPLICANT: Turner, Leah R.
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and Pr
; FILE REFERENCE: DEX-0432
; CURRENT APPLICATION NUMBER: US/10/517,696
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 60/389,327
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 171
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 149
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-517-696-149

Query Match      100.0%; Score 109; DB 6; Length 409;
Best Local Similarity 100.0%; Pred. No. 2.3e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HGVTSAPDTRPAPGSTAPPA 20
Db 145 HGVTSAPDTRPAPGSTAPPA 164

RESULT 14
US-10-517-696-141
; Sequence 141, Application US/10517696
; Publication No. US20060051759A1
; GENERAL INFORMATION:
; APPLICANT: diadexus, Inc.
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto A.
; APPLICANT: Turner, Leah R.
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and Pr
; FILE REFERENCE: DEX-0432
; CURRENT APPLICATION NUMBER: US/10/517,696
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 60/389,327
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 171
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 141
; LENGTH: 420
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-517-696-141

Query Match      100.0%; Score 109; DB 6; Length 420;
Best Local Similarity 100.0%; Pred. No. 2.4e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HGVTSAPDTRPAPGSTAPPA 20
Db 145 HGVTSAPDTRPAPGSTAPPA 164
```

```

Best Local Similarity 100.0%; Pred. No. 2.2e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HGVTSAPDTRPAPGSTAPPA 20
Db 145 HGVTSAPDTRPAPGSTAPPA 164

RESULT 13
US-10-517-696-149
; Sequence 149, Application US/10517696
; Publication No. US20060051759A1
; GENERAL INFORMATION:
; APPLICANT: diadexus, Inc.
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto A.
; APPLICANT: Turner, Leah R.
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and Pr
; FILE REFERENCE: DEX-0432
; CURRENT APPLICATION NUMBER: US/10/517,696
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 60/389,327
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 171
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 149
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-517-696-149

Query Match      100.0%; Score 109; DB 6; Length 409;
Best Local Similarity 100.0%; Pred. No. 2.3e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HGVTSAPDTRPAPGSTAPPA 20
Db 145 HGVTSAPDTRPAPGSTAPPA 164

RESULT 14
US-10-517-696-141
; Sequence 141, Application US/10517696
; Publication No. US20060051759A1
; GENERAL INFORMATION:
; APPLICANT: diadexus, Inc.
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto A.
; APPLICANT: Turner, Leah R.
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and Pr
; FILE REFERENCE: DEX-0432
; CURRENT APPLICATION NUMBER: US/10/517,696
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 60/389,327
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 171
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 141
; LENGTH: 420
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-517-696-141

Query Match      100.0%; Score 109; DB 6; Length 420;
Best Local Similarity 100.0%; Pred. No. 2.4e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HGVTSAPDTRPAPGSTAPPA 20
Db 145 HGVTSAPDTRPAPGSTAPPA 164
```

Db 81 HGVTSAPDTRPAPGSTAPPA 100

RESULT 15
US-10-517-696-121
; Sequence 121, Application US/10517696
; Publication NO. US20060051759A1
; GENERAL INFORMATION:
; APPLICANT: diaDexus, Inc.
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto A.
; APPLICANT: Turner, Lean R.
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and Peptides
; FILE REFERENCE: DEX-0432
; CURRENT APPLICATION NUMBER: US/10/517,696
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 60/389,327
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 171
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 121
; LENGTH: 463
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-517-696-121

Query Match 100.0%; Score 109; DB 6; Length 463;
Best Local Similarity 100.0%; Pred. NO. 2.6e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HGVTSAPDTRPAPGSTAPPA 20
Db : 124 HGVTSAPDTRPAPGSTAPPA 143

Search completed: April 14, 2006, 10:51:18
Job time : 17.0732 secs

This Page Blank (uspto)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2006, 10:46:57 ; Search time 17.9268 Seconds
(without alignments)
49.773 Million cell updates/sec

Title: US-09-606-910E-1
Perfect score: 113
Sequence: 1 AHGVTSAEDTRPAGSTAPPA 21

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 217505 seqs, 42489236 residues

Total number of hits satisfying chosen parameters: 217505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA New:*
1: /SIDSS/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /SIDSS/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
3: /SIDSS/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
4: /SIDSS/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
5: /SIDSS/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
6: /SIDSS/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
7: /SIDSS/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
8: /SIDSS/ptodata/2/pubpaa/US60_NEW_PUB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	113	100.0	21	7	US-11-088-546-8
2	113	100.0	109	6	US-10-517-696-115
3	113	100.0	316	6	US-10-517-696-126
4	113	100.0	325	6	US-10-517-696-148
5	113	100.0	336	6	US-10-517-696-153
6	113	100.0	350	6	US-10-517-696-125
7	113	100.0	372	6	US-10-517-696-129
8	113	100.0	379	6	US-10-517-696-150
9	113	100.0	396	6	US-10-517-696-147
10	113	100.0	398	6	US-10-517-696-136
11	113	100.0	409	6	US-10-517-696-149
12	113	100.0	420	6	US-10-517-696-141
13	113	100.0	463	6	US-10-517-696-121
14	113	100.0	475	6	US-10-501-035-309
15	113	100.0	475	6	US-10-517-696-117
16	113	100.0	475	6	US-10-515-872-10
17	113	100.0	483	6	US-10-517-696-132
18	113	100.0	484	6	US-10-517-696-120
19	113	100.0	485	6	US-10-517-696-142
20	113	100.0	492	6	US-10-517-696-131
21	113	100.0	515	6	US-10-517-696-114
22	113	100.0	515	7	US-11-233-510-20
23	113	100.0	517	6	US-10-517-696-144
24	113	100.0	524	6	US-10-517-696-113
25	113	100.0	524	6	US-10-517-696-122

26	113	100.0	614	6	US-10-517-696-128	Sequence 128, App
27	113	100.0	728	6	US-10-517-696-112	Sequence 112, App
28	113	100.0	1255	7	US-11-050-857-487	Sequence 487, App
29	113	100.0	1255	7	US-11-043-806-398	Sequence 398, App
30	113	100.0	1256	6	US-10-517-696-111	Sequence 111, App
31	109	96.5	20	7	US-11-033-365-138	Sequence 138, App
32	109	96.5	20	7	US-11-033-365-139	Sequence 139, App
33	109	96.5	24	7	US-11-033-365-140	Sequence 140, App
34	98	86.7	156	6	US-10-401-386B-44	Sequence 44, Appl
35	98	86.7	307	6	US-10-401-386B-31	Sequence 31, Appl
36	94	83.2	461	6	US-10-515-872-14	Sequence 14, Appl
37	87	77.0	174	6	US-10-517-696-116	Sequence 116, App
38	87	77.0	256	6	US-10-517-696-130	Sequence 130, App
39	87	77.0	435	6	US-10-517-696-123	Sequence 123, App
40	82	72.6	20	6	US-10-515-872-4	Sequence 4, Appl
41	78	69.0	20	6	US-10-618-481-50	Sequence 50, Appl
42	76	67.3	20	7	US-11-126-798-33	Sequence 33, Appl
43	76	67.3	23	7	US-11-126-798-34	Sequence 34, Appl
44	74	65.5	20	6	US-10-515-872-7	Sequence 7, Appl
45	63	55.8	20	6	US-10-515-872-8	Sequence 8, Appl

ALIGNMENTS

RESULT 1
US-11-088-546-8
; Sequence 8, Application US/11088546
; Publication No. US20060063736A1
; GENERAL INFORMATION:
; APPLICANT: Carolyn R. Bertozzi
; TITLE OF INVENTION: Compositions and methods for inhibiting
; TITLE OF INVENTION: mucin-type O-linked glycosylation
; FILE REFERENCE: BERK-028
; CURRENT APPLICATION NUMBER: US/11/088,546
; PRIOR FILING DATE: 2005-03-23
; PRIOR APPLICATION NUMBER: 60/556,673
; FILING DATE: 2004-03-25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: ppGalNACT substrate
US-11-088-546-8

Query Match 100.0%; Score 113; DB 7; Length 21;
Best Local Similarity 100.0%; Pred. No. 3.7e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AHGVTSAEDTRPAGSTAPPA 21
|||||
DB 1 AHGVTSAEDTRPAGSTAPPA 21

RESULT 2
US-10-517-696-115
; Sequence 115, Application US/10517696
; Publication No. US20060051759A1
; GENERAL INFORMATION:
; APPLICANT: diaDexus, Inc.
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto A.
; APPLICANT: Turner, Leah R.
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and P
; FILE REFERENCE: DEX-0432
; CURRENT APPLICATION NUMBER: US/10/517,696
; CURRENT FILING DATE: 2004-12-13

; PRIOR APPLICATION NUMBER: US 60/389,327
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 171
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 115
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-517-696-115

Query Match 100.0%; Score 113; DB 6; Length 109;
Best Local Similarity 100.0%; Pred. No. 2e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AHGVTSA PDTRPAGSTAPPA 21
|||||
Db 39 AHGVTSA PDTRPAGSTAPPA 59

RESULT 3
US-10-517-696-126
; Sequence 126, Application US/10517696
; Publication No. US20060051759A1
; GENERAL INFORMATION:
; APPLICANT: diaDexus, Inc.
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto A.
; APPLICANT: Turner, Leah R.
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenchua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and Peptides
; FILE REFERENCE: DEX-0432
; CURRENT APPLICATION NUMBER: US/10/517,696
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 60/389,327
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 171
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 126
; LENGTH: 316
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-517-696-126

Query Match 100.0%; Score 113; DB 6; Length 316;
Best Local Similarity 100.0%; Pred. No. 6e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AHGVTSA PDTRPAGSTAPPA 21
|||||
Db 144 AHGVTSA PDTRPAGSTAPPA 164

RESULT 4
US-10-517-696-148
; Sequence 148, Application US/10517696
; Publication No. US20060051759A1
; GENERAL INFORMATION:
; APPLICANT: diaDexus, Inc.
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto A.
; APPLICANT: Turner, Leah R.
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenchua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and Peptides
; FILE REFERENCE: DEX-0432
; CURRENT APPLICATION NUMBER: US/10/517,696
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 60/389,327
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 171
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 148

; LENGTH: 325
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-517-696-148

Query Match 100.0%; Score 113; DB 6; Length 325;
Best Local Similarity 100.0%; Pred. No. 6.2e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AHGVTSA PDTRPAGSTAPPA 21
|||||
Db 144 AHGVTSA PDTRPAGSTAPPA 164

RESULT 5
US-10-517-696-153
; Sequence 153, Application US/10517696
; Publication No. US20060051759A1
; GENERAL INFORMATION:
; APPLICANT: diaDexus, Inc.
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto A.
; APPLICANT: Turner, Leah R.
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenchua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and Peptides
; FILE REFERENCE: DEX-0432
; CURRENT APPLICATION NUMBER: US/10/517,696
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 60/389,327
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 171
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 153
; LENGTH: 336
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-517-696-153

Query Match 100.0%; Score 113; DB 6; Length 336;
Best Local Similarity 100.0%; Pred. No. 6.4e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AHGVTSA PDTRPAGSTAPPA 21
|||||
Db 144 AHGVTSA PDTRPAGSTAPPA 164

RESULT 6
US-10-517-696-125
; Sequence 125, Application US/10517696
; Publication No. US20060051759A1
; GENERAL INFORMATION:
; APPLICANT: diaDexus, Inc.
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto A.
; APPLICANT: Turner, Leah R.
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenchua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and Peptides
; FILE REFERENCE: DEX-0432
; CURRENT APPLICATION NUMBER: US/10/517,696
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 60/389,327
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 171
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 125
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-517-696-125

Query Match 100.0%; Score 113; DB 6; Length 350;
Best Local Similarity 100.0%; Pred. No. 6.7e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AHGVTSAADTRPAGSTAPPA 21
Db 144 AHGVTSAADTRPAGSTAPPA 164

RESULT 7
US-10-517-696-129
; Sequence 129, Application US/10517696
; Publication No. US20060051759A1
; GENERAL INFORMATION:
; APPLICANT: diadexus, Inc.
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto A.
; APPLICANT: Turner, Leah R.
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and P
; CURRENT APPLICATION NUMBER: US/10/517,696
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 60/389,327
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 171
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 129
; LENGTH: 372
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-517-696-129

Query Match 100.0%; Score 113; DB 6; Length 372;
Best Local Similarity 100.0%; Pred. No. 7.1e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AHGVTSAADTRPAGSTAPPA 21
Db 144 AHGVTSAADTRPAGSTAPPA 164

RESULT 8
US-10-517-696-150
; Sequence 150, Application US/10517696
; Publication No. US20060051759A1
; GENERAL INFORMATION:
; APPLICANT: diadexus, Inc.
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto A.
; APPLICANT: Turner, Leah R.
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and P
; CURRENT APPLICATION NUMBER: US/10/517,696
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 60/389,327
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 171
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 150
; LENGTH: 379
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-517-696-150

Query Match 100.0%; Score 113; DB 6; Length 379;
Best Local Similarity 100.0%; Pred. No. 7.3e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AHGVTSAADTRPAGSTAPPA 21

Db 144 AHGVTSAADTRPAGSTAPPA 164

RESULT 9
US-10-517-696-147
; Sequence 147, Application US/10517696
; Publication No. US20060051759A1
; GENERAL INFORMATION:
; APPLICANT: diadexus, Inc.
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto A.
; APPLICANT: Turner, Leah R.
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and P
; CURRENT APPLICATION NUMBER: US/10/517,696
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 60/389,327
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 171
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 147
; LENGTH: 396
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-517-696-147

Query Match 100.0%; Score 113; DB 6; Length 396;
Best Local Similarity 100.0%; Pred. No. 7.6e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AHGVTSAADTRPAGSTAPPA 21
Db 144 AHGVTSAADTRPAGSTAPPA 164

RESULT 10
US-10-517-696-136
; Sequence 136, Application US/10517696
; Publication No. US20060051759A1
; GENERAL INFORMATION:
; APPLICANT: diadexus, Inc.
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto A.
; APPLICANT: Turner, Leah R.
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and P
; CURRENT APPLICATION NUMBER: US/10/517,696
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 60/389,327
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 171
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 136
; LENGTH: 398
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-517-696-136

Query Match 100.0%; Score 113; DB 6; Length 398;
Best Local Similarity 100.0%; Pred. No. 7.6e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AHGVTSAADTRPAGSTAPPA 21
Db 144 AHGVTSAADTRPAGSTAPPA 164

RESULT 11

US-10-517-696-149
; Sequence 149, Application US/10517696
; Publication No. US20060051759A1
; GENERAL INFORMATION:
; APPLICANT: diaDexus, Inc.
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto A.
; APPLICANT: Turner, Leah R.
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and Peptides
; FILE REFERENCE: DEX-0432
; CURRENT APPLICATION NUMBER: US/10/517,696
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 60/389,327
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 171
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 149
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-517-696-149

Query Match 100.0%; Score 113; DB 6; Length 409;
Best Local Similarity 100.0%; Pred. No. 7.9e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AHGVTSA PDTRPAGSTAPPA 21
|||||
DB 144 AHGVTSA PDTRPAGSTAPPA 164

RESULT 12
; Sequence 141, Application US/10517696
; Publication No. US20060051759A1
; GENERAL INFORMATION:
; APPLICANT: diaDexus, Inc.
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto A.
; APPLICANT: Turner, Leah R.
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and Peptides
; FILE REFERENCE: DEX-0432
; CURRENT APPLICATION NUMBER: US/10/517,696
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 60/389,327
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 171
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 141
; LENGTH: 420
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-517-696-141

Query Match 100.0%; Score 113; DB 6; Length 420;
Best Local Similarity 100.0%; Pred. No. 8.1e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AHGVTSA PDTRPAGSTAPPA 21
|||||
DB 80 AHGVTSA PDTRPAGSTAPPA 100

RESULT 13
US-10-517-696-121
; Sequence 121, Application US/10517696
; Publication No. US20060051759A1
; GENERAL INFORMATION:
; APPLICANT: diaDexus, Inc.

; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto A.
; APPLICANT: Turner, Leah R.
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and Peptides
; FILE REFERENCE: DEX-0432
; CURRENT APPLICATION NUMBER: US/10/517,696
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 60/389,327
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 171
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 121
; LENGTH: 463
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-517-696-121

Query Match 100.0%; Score 113; DB 6; Length 463;
Best Local Similarity 100.0%; Pred. No. 8.9e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AHGVTSA PDTRPAGSTAPPA 21
|||||
DB 123 AHGVTSA PDTRPAGSTAPPA 143

RESULT 14
US-10-501-035-309
; Sequence 309, Application US/10501035
; Publication No. US20060046249A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: IDENTIFICATION OF POLYNUCLEOTIDES AND POLYPEPTIDE FOR PREDICTING
; TITLE OF INVENTION: ACTIVITY OF COMPOUNDS THAT INTERACT WITH PROTEIN TYROSINE KINASE
; TITLE OF INVENTION: AND/OR PROTEIN TYROSINE KINASE PATHWAYS
; FILE REFERENCE: D0185 PCT
; CURRENT APPLICATION NUMBER: US/10/501,035
; CURRENT FILING DATE: 2004-07-09
; PRIOR APPLICATION NUMBER: US 60/350,061
; PRIOR FILING DATE: 2002-01-18
; NUMBER OF SEQ ID NOS: 795
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 309
; LENGTH: 475
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-501-035-309

Query Match 100.0%; Score 113; DB 6; Length 475;
Best Local Similarity 100.0%; Pred. No. 9.2e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AHGVTSA PDTRPAGSTAPPA 21
|||||
DB 135 AHGVTSA PDTRPAGSTAPPA 155

RESULT 15
US-10-517-696-117
; Sequence 117, Application US/10517696
; Publication No. US20060051759A1
; GENERAL INFORMATION:
; APPLICANT: diaDexus, Inc.
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto A.
; APPLICANT: Turner, Leah R.
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and Peptides
; FILE REFERENCE: DEX-0432
; CURRENT APPLICATION NUMBER: US/10/517,696

Tue Apr 18 08:15:24 2006

```
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 60/389,327
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 171
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 117
; LENGTH: 475
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-517-696-117

Query Match      100.0%; Score 113; DB 6; Length 475;
Best Local Similarity 100.0%; Pred. No. 9.2e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 AHGVTSA PDTTPAPGSTAPPA 21
Db      135 AHGVTSA PDTTPAPGSTAPPA 155
```

Search completed: April 14, 2006, 10:51:18
Job time : 18.9268 secs

It is Page Blank (uspto)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2006, 10:45:37 ; Search time 118.829 Seconds
(without alignments)
73.841 Million cell updates/sec

Title: US-09-606-910E-1

Perfect score: 113

Sequence: 1 AHGVSAPDTRPAGSTAPPA 21

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA Main:
1: /cgm2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgm2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
3: /cgm2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
4: /cgm2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
5: /cgm2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
6: /cgm2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	113	100.0	25	3	US-09-881-339-3
2	113	100.0	25	4	US-10-417-633-3
3	113	100.0	40	3	US-09-996-069-9
4	113	100.0	40	5	US-10-635-211-4
5	113	100.0	43	4	US-10-106-876-14
6	113	100.0	46	3	US-09-996-069-5
7	113	100.0	46	4	US-10-106-876-1
8	113	100.0	100	3	US-09-965-131-6
9	113	100.0	159	5	US-10-635-211-9
10	113	100.0	475	4	US-10-417-312-1
11	113	100.0	475	5	US-10-696-639-59
12	113	100.0	475	5	US-10-756-149-5038
13	113	100.0	475	6	US-11-055-113-67
14	113	100.0	495	6	US-11-055-119-2
15	113	100.0	508	4	US-10-057-136-20
16	113	100.0	515	4	US-10-097-340-212
17	113	100.0	515	4	US-10-171-311-156
18	113	100.0	515	4	US-10-612-090-19
19	113	100.0	515	6	US-11-050-926-212
20	113	100.0	586	5	US-10-635-211-2
21	113	100.0	1255	3	US-09-996-069-10
22	113	100.0	1255	4	US-10-171-311-158
23	113	100.0	1255	4	US-10-177-293-311
24	113	100.0	1255	4	US-10-734-564-120
25	113	100.0	1255	5	US-10-473-484-2
26	113	100.0	1255	5	US-10-997-055-2
27	113	100.0	1255	6	US-11-037-713-20

28 113 100.0 1255 6 US-11-009-533-2 Sequence 2, Appli
29 109 96.5 20 4 US-10-292-896-1 Sequence 1, Appli
30 109 96.5 20 5 US-10-705-401-1 Sequence 1, Appli
31 109 96.5 24 4 US-10-292-896-2 Sequence 2, Appli
32 109 96.5 24 5 US-10-705-401-2 Sequence 2, Appli
33 109 96.5 26 3 US-09-815-346-5 Sequence 5, Appli
34 109 96.5 26 4 US-10-106-876-8 Sequence 8, Appli
35 109 96.5 26 4 US-10-106-876-20 Sequence 20, Appli
36 109 96.5 27 3 US-09-815-346-4 Sequence 4, Appli
37 109 96.5 27 3 US-09-870-691-1 Sequence 1, Appli
38 109 96.5 27 3 US-09-999-191-1 Sequence 1, Appli
39 109 96.5 27 4 US-10-106-876-7 Sequence 7, Appli
40 109 96.5 27 4 US-10-106-876-9 Sequence 9, Appli
41 109 96.5 27 4 US-10-106-876-12 Sequence 12, Appli
42 109 96.5 28 4 US-10-106-876-13 Sequence 13, Appli
43 102.5 90.7 49 4 US-10-471-607-14 Sequence 14, Appli
44 102 90.3 20 3 US-09-822-698A-7 Sequence 7, Appli
45 102 90.3 20 3 US-09-834-240-1 Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-881-339-3
; Sequence 3, Application US/09881339
; Publication No. US20030138860A1
; GENERAL INFORMATION:
; APPLICANT: Robertson, John F.R.
; APPLICANT: Graves, Catherine
; APPLICANT: Price, Michael R.
; TITLE OF INVENTION: Cancer Detection Methods and Reagents
; FILE REFERENCE: 02332-0030 (49409-264876)
; CURRENT APPLICATION NUMBER: US/09/881.339
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,886
; PRIOR FILING DATE: 2000-06-14
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 25
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic MUC1 peptide TAP2
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (9)..(9)
; OTHER INFORMATION: T is O-glycosylated with N-acetylgalactosamine
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (21)..(21)
; OTHER INFORMATION: T is O-glycosylated with N-acetylgalactosamine
US-09-881-339-3

Query Match 100.0%; Score 113; DB 3; Length 25;
Best Local Similarity 100.0%; Pred. No. 6.1e-07;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AHGVSAPDTRPAGSTAPPA 21
DB 5 AHGVSAPDTRPAGSTAPPA 25

RESULT 2
US-10-417-633-3
; Sequence 3, Application US/10417633
; Publication No. US20030232399A1
; GENERAL INFORMATION:
; APPLICANT: Robertson, John F.R.
; APPLICANT: Graves, Catherine
; APPLICANT: Price, Michael R.
; TITLE OF INVENTION: Cancer Detection Methods and Reagents

```
; FILE REFERENCE: 52623-0031 52623-284706
; CURRENT APPLICATION NUMBER: US/10/417,633
; CURRENT FILING DATE: 2003-04-16
; PRIOR APPLICATION NUMBER: US 60/211,886
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 09/881,339
; PRIOR FILING DATE: 2001-06-14
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 25
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic MUC1 peptide TAP2
; NAME/KEY: MISC FEATURE
; LOCATION: (9)..(9)
; OTHER INFORMATION: T is O-glycosylated with N-acetylgalactosamine
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (21)..(21)
; OTHER INFORMATION: T is O-glycosylated with N-acetylgalactosamine
US-10-417-633-3
```

```
Query Match 100.0%; Score 113; DB 4; Length 25;
Best Local Similarity 100.0%; Pred. No. 6.1e-07;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 AHGVSAPDTRPAGSTAPPA 21
| | | | | | | | | | | | | | | | | | | | |
DB 5 AHGVSAPDTRPAGSTAPPA 25
```

RESULT 3

```
US-09-996-069-9
; Sequence 9, Application US/09996069
; Publication No. US20030036199A1
; GENERAL INFORMATION:
; APPLICANT: Bamdad, Cynthia
; APPLICANT: Bamdad, R. Shoshana
; TITLE OF INVENTION: DIAGNOSTIC TUMOR MARKERS, DRUG SCREENING FOR TUMORIGENESIS INHIBIT
; TITLE OF INVENTION: AND COMPOSITIONS AND METHODS FOR TREATMENT OF CANCER
; FILE REFERENCE: M01015/70071
; CURRENT APPLICATION NUMBER: US/09/996,069
; CURRENT FILING DATE: 2001-11-27
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 40
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-996-069-9
```

```
Query Match 100.0%; Score 113; DB 3; Length 40;
Best Local Similarity 100.0%; Pred. No. 9.7e-07;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 AHGVSAPDTRPAGSTAPPA 21
| | | | | | | | | | | | | | | | | | | | |
DB 14 AHGVSAPDTRPAGSTAPPA 34
```

RESULT 4

```
US-10-635-211-4
; Sequence 4, Application US/10635211
; Publication No. US20050031649A1
; GENERAL INFORMATION:
; APPLICANT: Beijing HYD-VAX Biotechnology Co. Ltd
; TITLE OF INVENTION: A recombinant fusion protein comprising BCG heat shock protein 65
; TITLE OF INVENTION: and the epitope of MUC1
; FILE REFERENCE: FP03012US
; CURRENT APPLICATION NUMBER: US/10/635,211
```

```
; CURRENT FILING DATE: 2003-08-06
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4
; LENGTH: 40
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-635-211-4
```

```
Query Match 100.0%; Score 113; DB 5; Length 40;
Best Local Similarity 100.0%; Pred. No. 9.7e-07;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 AHGVSAPDTRPAGSTAPPA 21
| | | | | | | | | | | | | | | | | | | | |
DB 7 AHGVSAPDTRPAGSTAPPA 27
```

RESULT 5

```
US-10-106-876-14
; Sequence 14, Application US/10106876
; Publication No. US20030157160A1
; GENERAL INFORMATION:
; APPLICANT: BUDZYNSKI, WLADYSLAW A.
; APPLICANT: KOGANTY, R. RAO
; APPLICANT: KRANTZ, MARK J.
; APPLICANT: LONGENECKER, B. MICHAEL
; TITLE OF INVENTION: VACCINE FOR MODULATING BETWEEN T1 AND T2 IMMUNE
; TITLE OF INVENTION: RESPONSES
; FILE REFERENCE: 042881-0176
; CURRENT APPLICATION NUMBER: US/10/106,876
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: 60/278,698
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 43
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: peptide
US-10-106-876-14
```

```
Query Match 100.0%; Score 113; DB 4; Length 43;
Best Local Similarity 100.0%; Pred. No. 1e-06;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 AHGVSAPDTRPAGSTAPPA 21
| | | | | | | | | | | | | | | | | | | | |
DB 17 AHGVSAPDTRPAGSTAPPA 37
```

RESULT 6

```
US-09-996-069-5
; Sequence 5, Application US/09996069
; Publication No. US20030036199A1
; GENERAL INFORMATION:
; APPLICANT: Bamdad, Cynthia
; APPLICANT: Bamdad, R. Shoshana
; TITLE OF INVENTION: DIAGNOSTIC TUMOR MARKERS, DRUG SCREENING FOR TUMORIGENESIS INHIBIT
; TITLE OF INVENTION: AND COMPOSITIONS AND METHODS FOR TREATMENT OF CANCER
; FILE REFERENCE: M01015/70071
; CURRENT APPLICATION NUMBER: US/09/996,069
; CURRENT FILING DATE: 2001-11-27
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 46
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-996-069-5
```

Query Match 100.0%; Score 113; DB 3; Length 46;
Best Local Similarity 100.0%; Pred. No. 1.1e-06;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AHGVTSA PDTRPAGSTAPPA 21
|||||
DB 14 AHGVTSA PDTRPAGSTAPPA 34
|||||

RESULT 7
US-10-106-876-1
; Sequence 1, Application US/10106876
; Publication No. US20030157160A1
; GENERAL INFORMATION:
; APPLICANT: KOGANTY, R. RAO
; APPLICANT: KRANTZ, MARK J.
; APPLICANT: LONGENECKER, B. MICHAEL
; TITLE OF INVENTION: VACCINE FOR MODULATING BETWEEN T1 AND T2 IMMUNE
; FILE REFERENCE: 042881-0176
; CURRENT APPLICATION NUMBER: US/10/106,876
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: 60/278,698
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 46
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Peptide
US-10-106-876-1

Query Match 100.0%; Score 113; DB 4; Length 46;
Best Local Similarity 100.0%; Pred. No. 1.1e-06;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AHGVTSA PDTRPAGSTAPPA 21
|||||
DB 20 AHGVTSA PDTRPAGSTAPPA 40
|||||

RESULT 8
US-09-965-131-6
; Sequence 6, Application US/09965131
; Patent No. US20020160502A1
; GENERAL INFORMATION:
; APPLICANT: Chung, Maureen A.
; APPLICANT: Sharma, Surendra
; APPLICANT: Chang, Helena R.
; APPLICANT: O'Donnell, Mark A.
; TITLE OF INVENTION: RECOMBINANT BCG VACCINES FOR THE
; TITLE OF INVENTION: PREVENTION AND TREATMENT OF CANCER
; FILE REFERENCE: WII-014CP
; CURRENT APPLICATION NUMBER: US/09/965,131
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/235,455
; PRIOR FILING DATE: 2000-09-26
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 100
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-965-131-6

Query Match 100.0%; Score 113; DB 3; Length 100;
Best Local Similarity 100.0%; Pred. No. 2.4e-06;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AHGVTSA PDTRPAGSTAPPA 21
|||||
DB 19 AHGVTSA PDTRPAGSTAPPA 39
|||||

RESULT 9
US-10-635-211-9
; Sequence 9, Application US/10635211
; Publication No. US20050031649A1
; GENERAL INFORMATION:
; APPLICANT: Beijing HYDVAX Biotechnology Co. Ltd
; TITLE OF INVENTION: A recombinant fusion protein comprising BCG heat shock protein 65
; TITLE OF INVENTION: and the epitope of MUC1
; FILE REFERENCE: FP03012US
; CURRENT APPLICATION NUMBER: US/10/635,211
; CURRENT FILING DATE: 2003-08-06
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
; LENGTH: 159
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion Protein
US-10-635-211-9

Query Match 100.0%; Score 113; DB 5; Length 159;
Best Local Similarity 100.0%; Pred. No. 3.9e-06;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AHGVTSA PDTRPAGSTAPPA 21
|||||
DB 20 AHGVTSA PDTRPAGSTAPPA 40
|||||

RESULT 10
US-10-417-312-1
; Sequence 1, Application US/10417312
; Publication No. US20030235868A1
; GENERAL INFORMATION:
; APPLICANT: Dyax Corp
; TITLE OF INVENTION: Antibodies Specific for Mucin Polypeptide
; FILE REFERENCE: 2403/2002
; CURRENT APPLICATION NUMBER: US/10/417,312
; CURRENT FILING DATE: 2003-04-16
; PRIOR APPLICATION NUMBER: US 60/374,432
; PRIOR FILING DATE: 2002-04-22
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 475
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-417-312-1

Query Match 100.0%; Score 113; DB 4; Length 475;
Best Local Similarity 100.0%; Pred. No. 1.2e-05;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AHGVTSA PDTRPAGSTAPPA 21
|||||
DB 135 AHGVTSA PDTRPAGSTAPPA 155
|||||

RESULT 11
US-10-696-639-59
; Sequence 59, Application US/10696639
; Publication No. US20050037439A1
; GENERAL INFORMATION:
; APPLICANT: Pharmacia Corporation
; APPLICANT: Bournar, Maureen J.
; TITLE OF INVENTION: DIFFERENTIALLY EXPRESSED GENES INVOLVED IN CANCER, THE

```

; TITLE OF INVENTION: POLYPEPTIDES ENCODED THEREBY, AND METHODS OF USING THE SAME
; FILE REFERENCE: 01040/1
; CURRENT APPLICATION NUMBER: US/10/696.639
; CURRENT FILING DATE: 2003-10-29
; PRIOR APPLICATION NUMBER: 60/422,176
; PRIOR FILING DATE: 2002-10-29
; NUMBER OF SEQ ID NOS: 3114
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 59
; LENGTH: 475
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-696-639-59

Query Match      100.0%; Score 113; DB 5; Length 475;
Best Local Similarity 100.0%; Pred. No. 1.2e-05;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AHGVTSA PDTRPAGSTAPPA 21
Db 135 AHGVTSA PDTRPAGSTAPPA 155

RESULT 12
US-10-756-149-5038
; Sequence 5038, Application US/10756149
; Publication No. US20050181375A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natacha
; APPLICANT: Zlcnik, Albert
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSIS OF METASTATIC CANCER, COMPOSITIONS AND
; TITLE OF INVENTION: METHODS OF SCREENING FOR MODULATORS OF METASTATIC CANCER
; FILE REFERENCE: file
; CURRENT APPLICATION NUMBER: US/10/756.149
; CURRENT FILING DATE: 2004-01-12
; NUMBER OF SEQ ID NOS: 5818
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5038
; LENGTH: 475
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-756-149-5038

Query Match      100.0%; Score 113; DB 5; Length 475;
Best Local Similarity 100.0%; Pred. No. 1.2e-05;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AHGVTSA PDTRPAGSTAPPA 21
Db 135 AHGVTSA PDTRPAGSTAPPA 155

RESULT 13
US-11-055-119-67
; Sequence 67, Application US/11055119
; Publication No. US20050142640A1
; GENERAL INFORMATION:
; APPLICANT: Taylor-Papadimitriou, Joyce
; APPLICANT: Heukamp, Lukas Carl
; APPLICANT: Offringa, Rienk
; APPLICANT: Melief, Cornelis Johanna Maria
; APPLICANT: Thomas, Mireille
; TITLE OF INVENTION: MUC-1 derived peptides
; FILE REFERENCE: 029395-017
; CURRENT APPLICATION NUMBER: US/11/055.119
; CURRENT FILING DATE: 2005-02-11
; PRIOR APPLICATION NUMBER: US/09/658.621
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: US 60/187,215
; PRIOR FILING DATE: 2000-03-03
; PRIOR APPLICATION NUMBER: GB 9921242.5
; PRIOR FILING DATE: 1999-09-10
; PRIOR FILING DATE: 1999-09-08
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 495
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-055-119-2

Query Match      100.0%; Score 113; DB 6; Length 495;
Best Local Similarity 100.0%; Pred. No. 1.2e-05;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AHGVTSA PDTRPAGSTAPPA 21
Db 155 AHGVTSA PDTRPAGSTAPPA 175

RESULT 14
US-11-055-119-2
; Sequence 2, Application US/11055119
; Publication No. US20050142640A1
; GENERAL INFORMATION:
; APPLICANT: Taylor-Papadimitriou, Joyce
; APPLICANT: Heukamp, Lukas Carl
; APPLICANT: Offringa, Rienk
; APPLICANT: Melief, Cornelis Johanna Maria
; APPLICANT: Acres, Bruce
; APPLICANT: Thomas, Mireille
; TITLE OF INVENTION: MUC-1 derived peptides
; FILE REFERENCE: 029395-017
; CURRENT APPLICATION NUMBER: US/11/055.119
; CURRENT FILING DATE: 2005-02-11
; PRIOR APPLICATION NUMBER: US/09/658.621
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: US 60/187,215
; PRIOR FILING DATE: 2000-03-03
; PRIOR APPLICATION NUMBER: GB 9921242.5
; PRIOR FILING DATE: 1999-09-08
; PRIOR FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 495
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-055-119-2

Query Match      100.0%; Score 113; DB 6; Length 495;
Best Local Similarity 100.0%; Pred. No. 1.2e-05;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AHGVTSA PDTRPAGSTAPPA 21
Db 155 AHGVTSA PDTRPAGSTAPPA 175

RESULT 15
US-10-057-136-20
; Sequence 20, Application US/10057136
; Publication No. US20030021770A1
; GENERAL INFORMATION:
; APPLICANT: SCHLOW, JEFFREY
; APPLICANT: KANTOR, JUDITH
; APPLICANT: KUFE, DONALD
; APPLICANT: PANICALI, DENNIS
; APPLICANT: GRITZ, LINDA
; TITLE OF INVENTION: RECOMBINANT POX VIRUS FOR IMMUNIZATION AGAINST MUC1
; TITLE OF INVENTION: TUMOR-ASSOCIATED ANTIGEN
; FILE REFERENCE: 700953/47113C
; CURRENT APPLICATION NUMBER: US/10/057.136
```

; CURRENT FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: 09/366,670
; PRIOR FILING DATE: 1999-08-03
; PRIOR APPLICATION NUMBER: PCT/US98/03693
; PRIOR FILING DATE: 1998-02-24
; PRIOR APPLICATION NUMBER: 60/038,253
; PRIOR FILING DATE: 1997-02-24
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 508
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-057-136-20

Query Match 100.0%; Score 113; DB 4; Length 508;
Best Local Similarity 100.0%; Pred. No. 1.2e-05;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AHGVTSAPTDTPAPGSTAPPA 21
|||
Db 82 AHGVTSAPTDTPAPGSTAPPA 102
|||

Search completed: April 14, 2006, 10:50:37
Job time : 119.829 secs

This Page Blank (uspto)

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	113	100.0	25	2	US-09-217-306B-3		Sequence 3, Appli
2	113	100.0	40	1	US-08-099-354-1		Sequence 1, Appli
3	113	100.0	40	1	US-08-288-059-7		Sequence 7, Appli
4	109	96.5	24	2	US-08-737-896-5		Sequence 5, Appli
5	109	96.5	24	2	US-09-497-232-23		Sequence 23, Appli
6	109	96.5	24	4	PCT-US96-09951-5		Sequence 5, Appli
7	109	96.5	25	2	US-09-497-232-5		Sequence 5, Appli
8	109	96.5	30	2	US-08-737-896-6		Sequence 6, Appli
9	109	96.5	30	4	PCT-US96-09951-6		Sequence 6, Appli
10	102	90.3	20	1	US-08-833-807-8		Sequence 8, Appli
11	102	90.3	20	2	US-09-223-043-8		Sequence 8, Appli
12	102	90.3	20	2	US-09-291-351-1		Sequence 1, Appli
13	102	90.3	20	2	US-09-043-731-16		Sequence 16, Appli
14	102	90.3	20	2	US-09-593-870A-20		Sequence 20, Appli
15	102	90.3	20	2	US-09-834-240-1		Sequence 1, Appli
16	102	90.3	21	1	US-08-933-807-7		Sequence 7, Appli
17	102	90.3	21	2	US-09-423-043-7		Sequence 7, Appli
18	102	90.3	21	2	US-09-423-731-15		Sequence 15, Appli
19	102	90.3	21	2	US-09-593-870A-19		Sequence 19, Appli
20	102	90.3	32	2	US-09-593-870A-46		Sequence 46, Appli
21	101	89.4	20	2	US-09-339-944-6		Sequence 6, Appli
22	101	89.4	20	2	US-09-497-232-10		Sequence 10, Appli
23	101	89.4	20	2	US-08-651-265-6		Sequence 6, Appli
24	101	89.4	20	2	US-10-296-317-45		Sequence 45, Appli
25	101	89.4	24	2	US-09-217-306B-16		Sequence 16, Appli
26	101	89.4	24	2	US-09-217-306B-17		Sequence 17, Appli
27	101	89.4	25	1	US-08-288-059-28		Sequence 28, Appli

Tue Apr 18 08:15:23 2006

```

;
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/099,354
; FILING DATE: 30-JUL-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: SIRILLA, GEORGE M.
; REGISTRATION NUMBER: 18221
; REFERENCE/DOCKET NUMBER: 6137/202246
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-861-3536
; TELEFAX: 202-822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 40 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-099-354-1

Query Match 100.0%; Score 113; DB 1; Length 40;
Best Local Similarity 100.0%; Pred. No. 9.4e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AHGVTSA PDTRPAGSTAPPA 21
DB 15 AHGVTSA PDTRPAGSTAPPA 35

RESULT 3
US-08-288-059-7
; Sequence 7, Application US/08288059
; Patent No. 5827666
; GENERAL INFORMATION:
; APPLICANT: FINN, OLIVERA J.
; APPLICANT: FONTENOT, J. D.
; APPLICANT: MONTELARO, RONALD C.
; TITLE OF INVENTION: SYNTHETIC MULTIPLE TANDEM REPEAT MUCIN
; TITLE OF INVENTION: AND MUCIN-LIKE PEPTIDES, AND USES THEREOF
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/288,059
; FILING DATE: 08-AUG-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: CHAPIN, MARLANA K.
; REGISTRATION NUMBER: 35,843
; REFERENCE/DOCKET NUMBER: 61137/205204
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-861-3711
; TELEFAX: 202-822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 40 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-288-059-7
```

```

;
; LENGTH: 40 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-288-059-7

Query Match 100.0%; Score 113; DB 1; Length 40;
Best Local Similarity 100.0%; Pred. No. 9.4e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AHGVTSA PDTRPAGSTAPPA 21
DB 15 AHGVTSA PDTRPAGSTAPPA 35

RESULT 4
US-08-737-896-5
; Sequence 5, Application US/08737896
; Patent No. 6168804
; GENERAL INFORMATION:
; APPLICANT: Samuel, John
; APPLICANT: Kwon, Glen S.
; TITLE OF INVENTION: METHOD FOR ELICITING TH1-SPECIFIC
; TITLE OF INVENTION: IMMUNE RESPONSE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/737,896
; FILING DATE: 24-SEP-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/480,499
; FILING DATE: 07-JUN-1996
; APPLICATION NUMBER: PCT/US96/09551
; FILING DATE: 07-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07254/037001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 24 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; CLONE: SPQ-065 MUC1
; US-08-737-896-5

Query Match 96.5%; Score 109; DB 2; Length 24;
Best Local Similarity 100.0%; Pred. No. 1.8e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AHGVTSA PDTRPAGSTAPP 20
DB 5 AHGVTSA PDTRPAGSTAPP 24

RESULT 5
```

```

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/09951
; FILING DATE: 06-JUN-1996
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Learn, June M.
; REGISTRATION NUMBER: 31,238
; REFERENCE/DOCKET NUMBER: 07254/037W01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 678-5070
; TELEFAX: (619) 678-5099
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 24 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; CLONE: SPQ-065 MUC1
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..24
; PCT-US96-09951-5
;
; Query Match 96.5%; Score 109, DB 4, Length 24;
; Best Local Similarity 100.0%; Pred. No. 1.8e-07;
; Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 AHGVTSAPDTRPAGSTAPP 20
; DB 5 AHGVTSAPDTRPAGSTAPP 24
;
; RESULT 7
; US-09-497-232-5
; Sequence 5, Application US/09497232
; Patent No. 6600012
; GENERAL INFORMATION:
; APPLICANT: AGRAWAL, Babita
; KRANTZ, Mark J.
; REDDISH, Mark A.
; LONGENECKER, B. Michael
; TITLE OF INVENTION: METHOD FOR GENERATING ACTIVATED T-CELLS
; AND ANTIGEN-PULSED ANTIGEN-PRESENTING CELLS
;
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/497,232
; FILING DATE: 03-Feb-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/074,410
; FILING DATE: 08-MAY-1998
; APPLICATION NUMBER: US 60/045,949
; FILING DATE: 08-MAY-1997
; ATTORNEY/AGENT INFORMATION:

```

NAME: Saxe, Bernhard D.
REGISTRATION NUMBER: 28,665
REFERENCE/DOCKET NUMBER: 042881/0114
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 25 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-497-232-5

Query Match 96.5%; Score 109; DB 2; Length 25;
Best Local Similarity 100.0%; Pred. No. 1.8e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AHGVTSPDTRPAGSTAPP 20
Db 6 AHGVTSPDTRPAGSTAPP 25

RESULT 8
US-08-737-896-6
Sequence 6, Application US/08737896
Patent No. 616804
GENERAL INFORMATION:
APPLICANT: Samuel, John
APPLICANT: Kwon, Glen S.
TITLE OF INVENTION: METHOD FOR ELICITING TH1-SPECIFIC
TITLE OF INVENTION: IMMUNE RESPONSE
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/737,896
FILING DATE: 24-SEP-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/480,499
FILING DATE: 07-JUN-1996
APPLICATION NUMBER: PCT/US96/09551
FILING DATE: 07-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Haile, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07254/037001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 30 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
IMMEDIATE SOURCE:
CLONE: SP1-070 MUC1
US-08-737-896-6

Query Match 96.5%; Score 109; DB 2; Length 30;

Best Local Similarity 100.0%; Pred. No. 2.2e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AHGVTSPDTRPAGSTAPP 20
Db 11 AHGVTSPDTRPAGSTAPP 30

RESULT 9
PCT-US96-09951-6
Sequence 6, Application PC/TUS9609951
GENERAL INFORMATION:
APPLICANT: The Governors of the University of Alberta
TITLE OF INVENTION: A METHOD FOR ELICITING A TH1-SPECIFIC
TITLE OF INVENTION: IMMUNE RESPONSE
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: California
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/09951
FILING DATE: 06-JUN-1996
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Learn, June M.
REGISTRATION NUMBER: 31,238
REFERENCE/DOCKET NUMBER: 07254/037W01
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 678-5070
TELEFAX: (619) 678-5099
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 30 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
IMMEDIATE SOURCE:
CLONE: SP1-070 MUC1
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..30
PCT-US96-09951-6

Query Match 96.5%; Score 109; DB 4; Length 30;
Best Local Similarity 100.0%; Pred. No. 2.2e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AHGVTSPDTRPAGSTAPP 20
Db 11 AHGVTSPDTRPAGSTAPP 30

RESULT 10
US-08-833-807-8
Sequence 8, Application US/08833807
Patent No. 598952
GENERAL INFORMATION:
APPLICANT: McKenzie, Ian F.C.
APPLICANT: Apostolopoulos, Vasso
APPLICANT: Pietersz, Geoff A.
TITLE OF INVENTION: ANTIGENIC CARBOHYDRATE COMPOUNDS AND
TITLE OF INVENTION: THEIR USE IN IMMUNOTHERAPY
NUMBER OF SEQUENCES: 14

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Dann Dorfman Herrell and Skillman
;; STREET: Suite 720, 1601 Market Street
;; CITY: Philadelphia
;; STATE: Pennsylvania
;; COUNTRY: United States of America
;; ZIP: 19103-2307
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/833,807
;; FILING DATE:
;; CLASSIFICATION: 424
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/340,711
;; FILING DATE: 16-NOV-1994
;; APPLICATION NUMBER: AU PM3223
;; FILING DATE: 24-DEC-1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Hagan, Patrick J.
;; REGISTRATION NUMBER: 27,643
;; REFERENCE/DOCKET NUMBER: 530547/PAS/MKR
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (215)563-4100
;; TELEFAX: (215)563-4044
;; INFORMATION FOR SEQ ID NO: 8:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 20 amino acids
;; TYPE: amino acid
;; STRANDEDNESS:
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
;; US-08-833-807-8
Query Match 90.3%; Score 102; DB 1; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.1e-06;
Matches 19; Conservative 0; Mismatches 0; Indels 0;
Oy 1 AHGVSAPDTRPAPGSTAP 19
Db 2 AHGVSAPDTRPAPGSTAP 20
RESULT 11
US-09-223-043-8
; Sequence 8, Application US/09223043
; Patent No. 617256
; GENERAL INFORMATION:
; APPLICANT: McKenzie, Ian F.C.
; APPLICANT: Apostolopoulos, Vassio
; APPLICANT: Pietersz, Geoff A.
; TITLE OF INVENTION: ANTIGENIC CARBOHYDRATE COMPOUNDS AND
; TITLE OF INVENTION: THEIR USE IN IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann Dorfman Herrell and Skillman
; STREET: Suite 720, 1601 Market Street
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: United States of America
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/223,043
; FILING DATE:
; CLASSIFICATION:

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/833,807
;; FILING DATE:
;; APPLICATION NUMBER: AU PM3223
;; FILING DATE: 24-DEC-1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Hagan, Patrick J.
;; REGISTRATION NUMBER: 27,643
;; REFERENCE/DOCKET NUMBER: 530547/PAS/MKR
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (215)563-4100
;; TELEFAX: (215)563-4044
;; INFORMATION FOR SEQ ID NO: 8:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 20 amino acids
;; TYPE: amino acid
;; STRANDEDNESS:
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
;; US-09-223-043-8
Query Match 90.3%; Score 102; DB 2; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.1e-06;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 AHGVSAPDTRPAPGSTAP 19
Db 2 AHGVSAPDTRPAPGSTAP 20
RESULT 12
US-09-291-351-1
; Sequence 1, Application US/09291351
; Patent No. 6281004
; GENERAL INFORMATION:
; APPLICANT: Bogen, Steven A.
; APPLICANT: Radcliffe, Gail E.
; TITLE OF INVENTION: Quality Control for Cytochemical Assays
; FILE REFERENCE: CYL98-03
; CURRENT APPLICATION NUMBER: US/09/291,351
; CURRENT FILING DATE: 1999-04-14
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Human
; US-09-291-351-1
Query Match 90.3%; Score 102; DB 2; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.1e-06;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 AHGVSAPDTRPAPGSTAP 19
Db 2 AHGVSAPDTRPAPGSTAP 20
RESULT 13
US-09-043-731-16
; Sequence 16, Application US/09043731A
; Patent No. 6344203
; GENERAL INFORMATION:
; APPLICANT: The Austin Research Institute
; TITLE OF INVENTION: Mimicking Peptides in Cancer Therapy
; FILE REFERENCE: CALA-200
; CURRENT APPLICATION NUMBER: US/09/043,731A
; CURRENT FILING DATE: 1998-06-23
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 20
; TYPE: PRT

```
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: single
; OTHER INFORMATION: stranded linear peptide
US-09-043-731-16

Query Match          90.3%; Score 102; DB 2; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.1e-06;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AHGVTSA PDTRPAPGSTAP 19
    |||||
Db 2 AHGVTSA PDTRPAPGSTAP 20
    |||||

RESULT 14
US-09-593-870A-20
; Sequence 20, Application US/09593870A
; Patent No. 6548643
; GENERAL INFORMATION:
; APPLICANT: McKenzie, Ian F.C.
; APPLICANT: Apostolopoulos, Vasso
; APPLICANT: Pietersz, Geoff Allan
; TITLE OF INVENTION: Antigen Carbohydrate Compounds and Their
; TITLE OF INVENTION: Use in Immunotherapy
; FILE REFERENCE: 2368-McKenzie
; CURRENT APPLICATION NUMBER: US/09/593,870A
; CURRENT FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: 09/223,043
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 20
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-593-870A-20

Query Match          90.3%; Score 102; DB 2; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.1e-06;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AHGVTSA PDTRPAPGSTAP 19
    |||||
Db 2 AHGVTSA PDTRPAPGSTAP 20
    |||||

RESULT 15
US-09-834-240-1
; Sequence 1, Application US/09834240
; Patent No. 6855490
; GENERAL INFORMATION:
; APPLICANT: Sompuram, Seshi R.
; APPLICANT: Ramanathan, Halasya
; TITLE OF INVENTION: Quality Control for Cytochemical Assays
; FILE REFERENCE: 1159.1008-005
; CURRENT APPLICATION NUMBER: US/09/834,240
; CURRENT FILING DATE: 2003-03-28
; PRIOR APPLICATION NUMBER: 09/549,855
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 09/291,351
; PRIOR FILING DATE: 1999-04-14
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-834-240-1

Query Match          90.3%; Score 102; DB 2; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.1e-06;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Oy 1 AHGVTSA PDTRPAPGSTAP 19
|||
Db 2 AHGVTSA PDTRPAPGSTAP 20
|||

Search completed: April 14, 2006, 10:31:39
Job time : 36.3415 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2006, 10:14:04 ; Search time 159.293 Seconds
(without alignments)
93.012 Million cell updates/sec

Title: US-09-606-910E-1
Perfect score: 113
Sequence: 1 AHGVTSAPDTRPAGSTAPPA 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	113	100.0	475	2 Q4VX32 HUMAN	Q4VX32 homo sapien
2	113	100.0	1255	1 MUC1 HUMAN	P15941 h mucin-1 p
3	97	85.8	256	2 Q7M4M7 HUMAN	Q7M4M7 homo sapien
4	91	80.5	475	1 MUC1 HYLLA	Q29435 hylobates l
5	85	75.2	553	2 Q9MZL1 MACMU	Q9MZL1 macaca mula
6	78	69.0	20	2 Q9UMI8 HUMAN	Q9UMI8 homo sapien
7	65	57.5	745	2 Q89X06 BRAJA	Q89X06 bradyrhizob
8	62	54.9	564	2 Q7F233 ORYSA	Q7F233 oryza sativ
9	62	54.9	1334	2 Q9RKR9 STRCO	Q9RKR9 streptomyce
10	61	54.0	516	2 Q9XVH5 STRPU	Q9XVH5 strongyloce
11	60	53.1	162	2 Q8MJW2 EQUHM	Q8MJW2 equus hemio
12	60	53.1	162	2 Q8MJW4 EQUAS	Q8MJW4 equus asinu
13	60	53.1	168	2 Q8MJV8 EQUZE	Q8MJV8 equus zebra
14	60	53.1	168	2 Q8MJV9 EQUGR	Q8MJV9 equus grevy
15	60	53.1	705	2 Q848C3 STRHY	Q848C3 streptomyce
16	59	52.2	360	2 Q4UW97 XANCP	Q4UW97 xanthomonas
17	59	52.2	360	2 Q8P7V2 XANCP	Q8P7V2 xanthomonas
18	59	52.2	889	2 Q7F2N5 STRCO	Q7F2N5 streptomyce
19	58	51.3	596	2 Q7S5L7 NEUCR	Q7S5L7 neurospora
20	57	50.4	428	2 Q22369 CAEBL	Q22369 caenorhabdi
21	56	49.6	267	2 Q9A346 CAUCR	Q9A346 caulobacter
22	56	49.6	569	2 Q9KG99 LISMO	Q9KG99 listeria mo
23	56	49.6	909	2 Q6MP91 BDEBA	Q6MP91 bdellovibri
24	56	49.6	2706	2 O15870 PLAPA	O15870 plasmodium
25	55	48.7	310	1 PURU MYCBO	P0A5T7 mycobacteri
26	55	48.7	310	1 PURU MYCTU	P0A5T6 mycobacteri
27	55	48.7	329	2 Q9SMI5 MAIZE	Q9SMI5 zea mays (m
28	55	48.7	400	2 Q73VH2 MYCPA	Q73VH2 mycobacteri
29	55	48.7	429	2 Q6IPB3 CABBR	Q6IPB3 caenorhabdi
30	55	48.7	569	2 Q71X10 LISMF	Q71X10 listeria mo
31	55	48.7	700	2 Q521Z4 MAGGR	Q521Z4 magnaporthe

32	55	48.7	1366	2 Q4RSK8 TETNG	Q4RSK8 tetraodon n
33	55	48.7	2331	2 Q59EG0 HUMAN	Q59EG0 homo sapien
34	55	48.7	4391	1 PGBM HUMAN	P98160 homo sapien
35	55	48.7	4391	2 Q5VU27 HUMAN	Q5VU27 homo sapien
36	54	47.8	105	2 Q5JWB5 HUMAN	Q5JWB5 homo sapien
37	54	47.8	291	2 Q7S4Q2 NEUCR	Q7S4Q2 neurospora
38	54	47.8	299	2 Q4NR67 XENTR	Q4NR67 anseromyxob
39	54	47.8	363	2 Q6F2X3 XENTR	Q6F2X3 xenopus tro
40	54	47.8	440	2 Q9EX20 STRCO	Q9EX20 streptomyce
41	54	47.8	484	1 SUFU HUMAN	Q9UMX1 homo sapien
42	54	47.8	508	2 Q8CD55 MOUSE	Q8CD55 mus musculu
43	54	47.8	526	2 Q4R7F4 MACFA	Q4R7F4 macaca fasc
44	54	47.8	529	1 FGRL1 RAT	Q7LQM3 rattus norv
45	54	47.8	529	2 Q4V8P8 RAT	Q4V8P8 rattus norv

ALIGNMENTS

RESULT 1
Q4VX32 HUMAN
ID Q4VX32 HUMAN PRELIMINARY; PRT; 475 AA.
AC Q4VX32;
DT 13-SEP-2005 (Tremblrel. 31, Created)
DT 13-SEP-2005 (Tremblrel. 31, Last sequence update)
DE Mucin 1, transmembrane.
GN Name=MUC1; ORFNames=RP11-263K19.2-001;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Gluthero R.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL713999; CAI95078.1; -; Genomic_DNA.
KW Transmembrane.
SQ SEQUENCE 475 AA; 49224 MW; CB64D56690818614 CRC64;

Query Match 100.0%; Score 113; DB 2; Length 475;
Best Local Similarity 100.0%; Pred. No. 8.3e-06;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AHGVTSAPDTRPAGSTAPPA 21
|||||
Db 135 AHGVTSAPDTRPAGSTAPPA 155
|||||

RESULT 2
MUC1 HUMAN
ID MUC1 HUMAN STANDARD; PRT; 1255 AA.
AC P15941; P13931; P15942; P17626; Q14128; Q14876; Q16437; Q16442;
AC Q16615; Q9BXA4; Q9UE75; Q9UE76; Q9UQ11; Q9Y4J2;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Mucin-1 precursor (MUC-1) (Polymorphic epithelial mucin) (PEM) (PEMT)
DE (Episialin) (Tumor-associated mucin) (Carcinoma-associated mucin)
DE (Tumor-associated epithelial membrane antigen) (EMA) (H23AG) (Peanut-
DE reactive urinary mucin) (PUM) (Breast carcinoma-associated antigen
DE DF3) (CD227 antigen).
GN Name=MUC1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE (ISOFORM 1).
RC TISSUE=Pancreas;
RX MEDLINE=90368716; PubMed=2394722;

RA Ian M.S., Batra S.K., Qi W.-N., Metzgar R.S., Hollingsworth M.A.;
RT "Cloning and sequencing of a human pancreatic tumor mucin cDNA.";
RL J. Biol. Chem. 265:15294-15299(1990).
RN [2]

RP MEDLINE=90202794; PubMed=2318925;
RX Lightenberg M.J.L., Vos H.L., Gennissen A.M.C., Hilkens J.;
RA "Epistatin, a carcinoma-associated mucin, is generated by a
RT polymorphic gene encoding splice variants with alternative amino
RT termini.";
RL J. Biol. Chem. 265:5573-5578(1990).
RN [3]

RP NUCLEOTIDE SEQUENCE (ISOFORM 1).
RC TISSUE=Mammary carcinoma;
RX MEDLINE=90368715; PubMed=1697589;
RA Gendler S.J., Lancaster C.A., Taylor-Papadimitriou J., Duhig T.,
RA Peat N., Burchell J., Pemberton L., Lalani E.-N., Wilson D.;
RT "Molecular cloning and expression of human tumor-associated
RT polymorphic epithelial mucin.";
RL J. Biol. Chem. 265:15286-15293(1990).
RN [4]

RP NUCLEOTIDE SEQUENCE (ISOFORM 1).
RX MEDLINE=91097524; PubMed=2268309;
RA Lancaster C.A., Peat N., Duhig T., Wilson D., Taylor-Papadimitriou J.,
RA Gendler S.J.;
RT "Structure and expression of the human polymorphic epithelial mucin
RT gene: an expressed VNTR unit.";
RL Biochem. Biophys. Res. Commun. 173:1019-1029(1990).
RN [5]

RP NUCLEOTIDE SEQUENCE (ISOFORM 5).
RC TISSUE=Mammary carcinoma;
RX MEDLINE=90276413; PubMed=2351132;
RA Wreschner D.H., Hareuveni M., Tsarfaty I., Smorodinsky N., Horev J.,
RA Zaretzky J., Kotkes P., Weiss M., Lathe R., Dion A., Keydar I.;
RT "Human epithelial tumor antigen cDNA sequences. Differential splicing
RT may generate multiple protein forms.";
RL Eur. J. Biochem. 189:463-473(1990).
RN [6]

RP TISSUE=Mammary carcinoma;
RX MEDLINE=90276414; PubMed=2112460;
RA Hareuveni M., Tsarfaty I., Zaretzky J., Kotkes P., Horev J.,
RA Zrihan S., Weiss M., Green S., Lathe R., Keydar I., Wreschner D.H.;
RT "A transcribed gene, containing a variable number of tandem repeats,
RT codes for a human epithelial tumor antigen. cDNA cloning, expression
RT of the transfected gene and over-expression in breast cancer tissue.";
RL Eur. J. Biochem. 189:475-486(1990).
RN [7]

RP NUCLEOTIDE SEQUENCE (ISOFORM 1).
RX MEDLINE=91033045; PubMed=1688329; DOI=10.1016/0378-1119(90)90242-J;
RA Tsarfaty I., Hareuveni M., Horev J., Zaretzky J., Weiss M.,
RA Jeltsch J.M., Garnier J.M., Lathe R., Keydar I., Wreschner D.H.;
RT "Isolation and characterization of an expressed hypervariable gene
RT coding for a breast-cancer-associated antigen.";
RL Gene 93:313-318(1990).
RN [8]

RP NUCLEOTIDE SEQUENCE (ISOFORM 7).
RX MEDLINE=95010060; PubMed=7925397;
RA Zrihan-Licht S., Vos H.L., Baruch A., Elroy-Stein O., Sagiv D.,
RA Keydar I., Hilkens J., Wreschner D.H.;
RT "Characterization and molecular cloning of a novel MUC1 protein,
RT devoid of tandem repeats, expressed in human breast cancer tissue.";
RL Eur. J. Biochem. 224:787-795(1994).
RN [9]

RP NUCLEOTIDE SEQUENCE (ISOFORMS 6; 7 AND 8).
RX MEDLINE=97355747; PubMed=9212228;
RA Oosterkamp H.M., Schelmer L., Stefanova M.C., Lloyd K.O.,
RA Finstad C.L.;
RT "Comparison of MUC-1 mucin expression in epithelial and non-epithelial
RT cancer cell lines and demonstration of a new short variant form (MUC-
RT 1/Z).";
RL Int. J. Cancer 72:87-94(1997).
RN [10]

RP NUCLEOTIDE SEQUENCE (ISOFORM 7), AND VARIANT MET-1117.
RA Zhang L.X., Li C.H.;
RT "Molecular cloning of an isoform of MUC1, MUC1/Y.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [11]

RP NUCLEOTIDE SEQUENCE (ISOFORM 9).
RC TISSUE=Epithelial cancer;
RX Zhang L.X., Li C.H., Sun L.Y., Yue W.;
RA "Cloning of a new potential secreted short variant form of MUC1 mucin
RT in epithelial cancer cell line.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [12]

RP NUCLEOTIDE SEQUENCE (GENOMIC DNA), AND VARIANTS MET-1117 AND ASN-1142.
RA Livingston R.J., Rieder M.J., Chung M.-W., Ritchie T.K., Olson A.N.,
RA Nguyen C.P., Nguyen D.A., Poel C.L., Robertson P.D., Schackwitz W.S.,
RA Sherwood J.K., Leithauser B.J., Nickerson D.A.;
RT "NIHES-SNPs, environmental genome project, NIHES ES15478, Department
RT of Genome Sciences, Seattle, WA [URL: <http://egp.gs.washington.edu>].";
RL Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.
RN [13]

RP PARTIAL NUCLEOTIDE SEQUENCE.
RX MEDLINE=88330762; PubMed=3417635;
RA Gendler S.J., Taylor-Papadimitriou J., Duhig T., Rothbard J.,
RA Burchell J.;
RT "A highly immunogenic region of a human polymorphic epithelial mucin
RT expressed by carcinomas is made up of tandem repeats.";
RL J. Biol. Chem. 263:12820-12823(1988).
RN [14]

RP NUCLEOTIDE SEQUENCE OF 1-160 (ISOFORM 2).
RX MEDLINE=90088473; PubMed=2597151;
RA Abe M., Siddiqui J., Kufe D.;
RT "Sequence analysis of the 5' region of the human DF3 breast carcinoma-
RT associated antigen gene.";
RL Biochem. Biophys. Res. Commun. 165:644-649(1989).
RN [15]

RP NUCLEOTIDE SEQUENCE OF 1-109 (ISOFORM 2).
RC TISSUE=Thyroid;
RX MEDLINE=96183746; PubMed=8608966;
RA Weiss M., Baruch A., Keydar I., Wreschner D.H.;
RT "Preoperative diagnosis of thyroid papillary carcinoma by reverse
RT transcriptase polymerase chain reaction of the MUC1 gene.";
RL Int. J. Cancer 66:55-59(1996).
RN [16]

RP NUCLEOTIDE SEQUENCE OF 1-89.
RC TISSUE=Lung;
RX MEDLINE=96181716; PubMed=8604237;
RA Yu C.J., Yang P.C., Shew J.Y., Hong T.M., Yang S.C., Lee Y.C.,
RA Lee L.N., Luh K.T., Wu C.W.;
RT "Mucin mRNA expression in lung adenocarcinoma cell lines and
RT tissues.";
RL Oncology 53:118-126(1996).
RN [17]

RP NUCLEOTIDE SEQUENCE OF 1-46 (ISOFORMS 3 AND 4).
RC TISSUE=Mammary carcinoma;
RA Buluwela L., Liu Q., Luqmani Y.A., Gomm J.J., Coombes R.C.;
RL Submitted (OCT-1992) to the EMBL/GenBank/DBJ databases.
RN [18]

RP CARBOHYDRATE-LINKAGE SITES WITHIN THE REPEAT.
RX MEDLINE=97460054; PubMed=9312074; DOI=10.1074/jbc.272.40.24780;
RA Mueller S., Goletz S., Packer N., Gooley A.A., Lawson A.M.,
RA Haniach F.-G.;
RT "Localization of O-glycosylation sites on glycopeptide fragments from
RT lactation-associated MUC1. All putative sites within the tandem repeat
RT are glycosylation targets in vivo.";
RL J. Biol. Chem. 272:24780-24793(1997).
RN [19]

RP CARBOHYDRATE-LINKAGE SITES WITHIN THE REPEAT.
RX MEDLINE=99303572; PubMed=10373415; DOI=10.1074/jbc.274.26.18165;
RA Mueller S., Alving K., Peter-Katalinic J., Zachara N., Gooley A.A.,
RA Haniach F.-G.;
RT "High density O-glycosylation on tandem repeat peptide from secretory
RT MUC1 of T47D breast cancer cells.";
RL J. Biol. Chem. 274:18165-18172(1999).
RN [20]


```

RN [20] POLYMORPHISM WITHIN THE REPEAT.
RX MEDLINE=21359366; PubMed=11350974; DOI=10.1074/jbc.M103187200;
RA Engelmann K., Baldus S.B., Hanisch F.-G.;
RT "Identification and topology of variant sequences within individual
RT repeat domains of the human epithelial tumor mucin MUC1.";
RL J. Biol. Chem. 276:27764-27769(2001).
RN [21]
RP CHARACTERIZATION OF ISOFORM Y, AND MUTAGENESIS OF ASP-1116.
RX MEDLINE=99211485; PubMed=10197628;
RA Baruch A., Hartmann M.-L., Yoeli M., Adereth Y., Greenstein S.,
RA Stadler Y., Skornik Y., Zaretsky J., Smorodinsky N.I., Keydar I.,
RA Wreschner D.H.;
RT "The breast cancer-associated MUC1 gene generates both a receptor and
RT its cognate binding protein.";
RL Cancer Res. 59:1552-1561(1999).
RN [22]
RP PARTIAL PROTEIN SEQUENCE, AND CHARACTERIZATION OF CLEAVAGE SITE.
RX MEDLINE=21240104; PubMed=11341784; DOI=10.1006/bbrc.2001.4775;
RA Parry S., Silverman H.S., McDermott K., Willis A., Hollingsworth M.A.,
RA Harris A.;
RT "Identification of MUC1 proteolytic cleavage sites in vivo.";
RL Biochem. Biophys. Res. Commun. 283:715-720(2001).
RN [23]
RP CHARACTERIZATION.
RX MEDLINE=21836452; PubMed=11847293; DOI=10.1110/ps.16502;
RA Wreschner D.H., McGuckin M.A., Williams S.J., Baruch A., Yoeli M.,
RA Ziv R., Okun L., Zaretsky J., Smorodinsky N., Keydar I., Neophytou P.,
RA Stacey M., Lin H.-H., Gordon S.;
RT "Generation of ligand-receptor alliances by 'SEA' module-mediated
RT cleavage of membrane-associated mucin proteins.";
RL Protein Sci. 11:698-706(2002).
RN [24]
Query Match 100.0%; Score 113; DB 1; Length 1255;
Best Local Similarity 100.0%; Pred. No. 2.3e-05; Indels 0; Gaps 0;
Matches 21; Conservative 0; Mismatches 0;

QY 1 AHGVTSA PDTRPAGSTAPPA 21
DB 135 AHGVTSA PDTRPAGSTAPPA 155

RESULT 3
QY 1 AHGVTSA PDTRPAGSTAPPA 21
DB 135 AHGVTSA PDTRPAGSTAPPA 155

ID Q7M4M7 HUMAN PRELIMINARY; PRT; 256 AA.
AC Q7M4M7.
DT 01-MAR-2004 (TREMBLrel. 26, Created)
DT 01-MAR-2004 (TREMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Tumor-associated antigen DF3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=90058554; PubMed=2582438;
RA Merlo G.R., Siddiqui J., Cropp C.S., Liscia D.S., Lidereau R.,
RA Callahan R., Kufe D.W.;
RT "Frequent alteration of the DF3 tumor-associated antigen gene in
RT primary human breast carcinomas.";
RL Cancer Res. 49:6966-6971(1989).
DR PIR; A60533; A60533.
DR Ensembl; ENSG00000185499; Homo sapiens.
DR GO; GO:0005198; P:structural molecule activity; IEA.
DR GO; GO:0019068; P:viral assembly; IEA.
DR InterPro; IPR000982; Matrix.
DR ProDom; PD000741; Matrix. 2.
SQ SEQUENCE 256 AA; 25766 MW; E0430FOBA7F30B04 CRC64;

Query Match 85.8%; Score 97; DB 2; Length 256;

```

```

Best Local Similarity 85.7%; Pred. No. 0.00041;
Matches 18; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 AHGVTSA PDTRPAGSTAPPA 21
DB 18 AHRVTSAPESRPAGSTAPPA 38

RESULT 4
MUC1_HYLLA
ID MUC1_HYLLA STANDARD; PRT; 475 AA.
AC Q29435;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Mucin-1 precursor (MUC-1).
GN Name=MUC1;
OS Hylobates lar (Common gibbon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
OC Hylobatidae; Hylobates.
OX NCBI_TaxID=9580;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=96351712; PubMed=8747930; DOI=10.1007/BF00292441;
RA Spicer A.P., Duhig T., Chilton B.S., Gendler S.J.;
RT "Analysis of mammalian MUC1 genes reveals potential functionally
RT important domains.";
RL Mamm. Genome 6:885-888(1995).
CC -!- FUNCTION: Direct or indirect interaction with actin cytoskeleton
CC (By similarity)
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- PTM: Highly O-glycosylated and probably also N-glycosylated.
CC -!- SIMILARITY: Contains 1 SEA domain.
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC EMBL; L41589; AAA69965.1; -; Genomic DNA.
CC EMBL; L41625; AAA69918.1; -; Genomic DNA.
CC EMBL; L41624; AAA69918.1; JOINED; Genomic DNA.
CC InterPro; IPR000082; SEA.
CC Pfam; PF01390; SEA; 1.
CC SMART; SM00200; SEA; 1.
CC PROSITE; PS50024; SEA; 1.
KW Actin-binding; Cytoskeleton; Glycoprotein; Repeat; Signal;
KW Transmembrane.
FT SIGNAL 1 23 Potential.
FT CHAIN 24 475 Mucin-1.
FT TOPO_DOM 24 380 Extracellular (Potential).
FT TRANSMEM 381 401 Potential.
FT TOPO_DOM 402 475 Cytoplasmic (Potential).
FT REPEAT 102 121 1.
FT REPEAT 122 141 2.
FT REPEAT 142 161 3.
FT REPEAT 162 181 4.
FT DOMAIN 254 371 SEA.
SQ SEQUENCE 475 AA; 49372 MW; D7A699D6D68C6622 CRC64;

Query Match 80.5%; Score 91; DB 1; Length 475;
Best Local Similarity 89.5%; Pred. No. 0.0043;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 HGVTSAPDTRPAGSTAPP 20
DB 136 HGVTSAPDTRPAGSTAPP 154

RESULT 5
Q9WZL1_MACMU

```

```

ID Q9WZL1_MACMU PRELIMINARY; PRT; 553 AA.
AC Q9WZL1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Mucin 1 (Fragment).
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
OC Cercopithecoidea; Cercopithecinae; Macaca.
OX NCBI_TaxID=9544;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=20330533; PubMed=10869775; DOI=10.1016/S0264-410X(00)00143-2;
RA Vaughan H.A., Ho D.W.M., Karanikas V., Sandrin M.S., McKenzie I.F.C.,
RA Pietersz G.A.;
RT "The immune response of mice and cynomolgus monkeys to macaque mucin
RT 1-nannan";
RL Vaccine 18:3297-3309(2000).
DR EMBL; AF176947; AAF82403.1; -; Genomic_DNA.
DR InterPro; IPR000082; SEA.
DR Pfam; PF01390; SEA; 1.
DR SMART; SM00200; SEA; 1.
DR PROSITE; PS50024; SEA; 1.
FT NON_TER 1
FT NON_TER 553
SQ SEQUENCE 553 AA; 55778 MW; 6D7E6DD2EE929318 CRC64;

Query Match 75.2%; Score 85; DB 2; Length 553;
Best Local Similarity 80.0%; Pred. No. 0.028;
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 HGVTSPDTRPAPGSTAPPA 21
DB 74 HNVTSAPDTSAPGSTGPPA 93

RESULT 6
ID Q9UM18_HUMAN PRELIMINARY; PRT; 20 AA.
AC Q9UM18;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Mucin (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=89235154; PubMed=2715633;
RA Xing P.X., Tjandra J.J., Reynolds K., McLaughlin P.J., Purcell D.F.J.,
RA McKenzie I.F.C.;
RT "Reactivity of anti-human milk fat globule antibodies with synthetic
RT peptides.";
RL J. Immunol. 142:3503-3509(1989).
DR EMBL; M26316; AAA36336.1; -; mRNA.
DR FIR; S10571; S10571.
FT NON_TER 1
FT NON_TER 20
SQ SEQUENCE 20 AA; 1887 MW; 5B3473EABFAFAD87 CRC64;

Query Match 69.0%; Score 78; DB 2; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.0065;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 PDTRPAPGSTAPPA 21
DB 1 PDTRPAPGSTAPPA 14

```

```

RESULT 7
Q89X06_BRAJA PRELIMINARY; PRT; 745 AA.
ID Q89X06;
AC Q89X06;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Blt0521 protein.
OS OrderedLocustNames=blt0521;
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiyama T.,
RA Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimo S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110.";
RL DNA Res. 9:189-197(2002).
DR EMBL; BA000040; BAC45786.1; -; Genomic_DNA.
DR HSSP; P07176; 1OAP.
DR GO; GO:0009279; C:outer membrane (sensu Gram-negative Bacteria); IEA.
DR GO; GO:0005199; F:structural constituent of cell wall; IEA.
DR InterPro; IPR006665; OmpA/MotB.
DR Pfam; PF003882; Pistil_extensin.
DR Pfam; PF00691; OmpA; 1.
DR PRINTS; PR01218; PSTEXTENSIN.
DR PRODOM; PD000930; OmpA/MotB; 1.
KW Complete proteome.
SQ SEQUENCE 745 AA; 74544 MW; 155EDFCC74DBC6D CRC64;

Query Match 57.5%; Score 65; DB 2; Length 745;
Best Local Similarity 65.0%; Pred. No. 11;
Matches 13; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 AHGVTSPDTRPAPGSTAPP 20
DB 272 APGATPAPTTPAPGGTATP 291

RESULT 8
ID Q7F233_ORYSA PRELIMINARY; PRT; 564 AA.
AC Q7F233;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE Putative PRL1-interacting factor G.
OS Name=OJ1354_H07.125; Synonyms=OJ1370_E02.114;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, BAC
RT clone:OJ1354_H07.125";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, BAC
RT clone:OJ1370_E02.114";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003755; BAC21341.1; -; Genomic DNA.
DR EMBL; AP003756; BAC30221.1; -; Genomic DNA.

```

Qy 3 GVTSA PDTRPAPGSTAPP 20
| | | | | | | | | |
Db 415 GTTPAPGTAPAPGSTTAP 432

```
DR GO; GO:0007186; P-G-protein coupled receptor protein signalin. . .; IEA.
DR InterPro; IPR002185; Dopad4_receptor.
DR PRINTS; PR00569; DOPAMINED4R.
KW Receptor.
FT NON_TER 162 162
FT NON_TER 162 162
SQ SEQUENCE 162 AA; 16608 MW; 0C742D20FC4A5C37 CRC64;

Query Match 53.1%; Score 60; DB 2; Length 162;
Best Local Similarity 64.7%; Pred. No. 9.5;
Matches 11; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 4 VTSAPDTRPAGSTAPP 20
  ||||| :|||
Db 87 VTPAPDATPPDATAPP 103

RESULT 12
QBMJW4_EQUAS
ID QBMJW4_EQUAS PRELIMINARY; PRT; 162 AA.
AC QBMJW4;
DT 01-OCT-2002 (TReMBLrel. 22, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE D4 dopamine receptor (Fragment).
GN Name=DRD4;
OS Equus asinus (Donkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9793;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Hasegawa T., Sato F., Ishida N.;
RT "Determination and variability of nucleotide sequences for D4 dopamine
  receptor genes (DRD4) in genus Equus.";
RL J. Equine Sci. 13:57-62(2002).
DR EMBL; AB080629; BAC10654.1; -; Genomic DNA.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0004952; F:dopamine receptor activity; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0007186; P-G-protein coupled receptor protein signalin. . .; IEA.
DR InterPro; IPR002185; Dopad4_receptor.
DR PRINTS; PR00569; DOPAMINED4R.
KW Receptor.
FT NON_TER 162 162
FT NON_TER 162 162
SQ SEQUENCE 162 AA; 16677 MW; 56183D31EA5B3A77 CRC64;

Query Match 53.1%; Score 60; DB 2; Length 162;
Best Local Similarity 64.7%; Pred. No. 9.5;
Matches 11; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 4 VTSAPDTRPAGSTAPP 20
  ||||| :|||
Db 87 VTPAPDATPPDATAPP 103

RESULT 13
QBMJV8_EQUZE
ID QBMJV8_EQUZE PRELIMINARY; PRT; 168 AA.
AC QBMJV8;
DT 01-OCT-2002 (TReMBLrel. 22, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE D4 dopamine receptor (Fragment).
GN Name=DRD4;
OS Equus zebra (Mountain zebra).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9791;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Hasegawa T., Sato F., Ishida N.;
```

```
RT "Determination and variability of nucleotide sequences for D4 dopamine
  receptor genes (DRD4) in genus Equus.";
RL J. Equine Sci. 13:57-62(2002).
DR EMBL; AB080635; BAC10660.1; -; Genomic DNA.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0004952; F:dopamine receptor activity; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0007186; P-G-protein coupled receptor protein signalin. . .; IEA.
DR InterPro; IPR002185; Dopad4_receptor.
DR PRINTS; PR00569; DOPAMINED4R.
KW Receptor.
FT NON_TER 168 168
FT NON_TER 168 168
SQ SEQUENCE 168 AA; 17316 MW; FA0D06B546D0B355 CRC64;

Query Match 53.1%; Score 60; DB 2; Length 168;
Best Local Similarity 64.7%; Pred. No. 9.9;
Matches 11; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 4 VTSAPDTRPAGSTAPP 20
  ||||| :|||
Db 87 VTPAPDATPPDATAPP 103

RESULT 14
QBMJV9_EQUGR
ID QBMJV9_EQUGR PRELIMINARY; PRT; 168 AA.
AC QBMJV9;
DT 01-OCT-2002 (TReMBLrel. 22, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE D4 dopamine receptor (Fragment).
GN Name=DRD4;
OS Equus grevyi (Grevy's zebra).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9792;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Hasegawa T., Sato F., Ishida N.;
RT "Determination and variability of nucleotide sequences for D4 dopamine
  receptor genes (DRD4) in genus Equus.";
RL J. Equine Sci. 13:57-62(2002).
DR EMBL; AB080634; BAC10659.1; -; Genomic DNA.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0004952; F:dopamine receptor activity; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0007186; P-G-protein coupled receptor protein signalin. . .; IEA.
DR InterPro; IPR002185; Dopad4_receptor.
DR PRINTS; PR00569; DOPAMINED4R.
KW Receptor.
FT NON_TER 168 168
FT NON_TER 168 168
SQ SEQUENCE 168 AA; 17258 MW; FB12046546D0B185 CRC64;

Query Match 53.1%; Score 60; DB 2; Length 168;
Best Local Similarity 64.7%; Pred. No. 9.9;
Matches 11; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 4 VTSAPDTRPAGSTAPP 20
  ||||| :|||
Db 87 VTPAPDATPPDATAPP 103

RESULT 15
Q848C3_STRHY
ID Q848C3_STRHY PRELIMINARY; PRT; 705 AA.
AC Q848C3;
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE Type I PKs (Fragment).
GN Name=gdnB;
```

OS Streptomyces hygroscopicus.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.

OX NCBI_TaxID=1912;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RA Wang Y., Gao Q.;

RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF521895; AAC61201.1; -; Genomic DNA.

DR GO: GO:0048037; F:cofactor binding; IEA.

DR InterPro: IPR009081; ACP like.

DR InterPro: IPR002198; ADH_short.

DR InterPro: IPR000183; Decarboxylase2.

DR InterPro: IPR006163; Phosphateth_bind.

DR InterPro: IPR006162; Ppantne S_.

DR Pfam: PF00106; adh_short; 1_.

DR PROSITE; PS00550; PP-Binding; 1.

DR PROSITE; PS00775; ACP_DOMAIN; 1.

DR PROSITE; PS00879; ODR_DC_2; UNKNOWN_1.

DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.

FT NON TER 1

SQ SEQUENCE 705 AA; 72867 MW; CEBD20B0EBC34990 CRC64;

Query Match 53.1%; Score 60; DB 2; Length 705;

Best Local Similarity 52.6%; Pred. No. 44;

Matches 10; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 3 GVTSPDTRPAGSTAPPA 21

Db 677 GATPSPGATPSPGATLPPA 695

Search completed: April 14, 2006, 10:29:33

Job time : 164.293 secs

This Page Blank (uspto)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2006, 10:24:48 ; Search time 24.5854 Seconds
(without alignments)
82.185 Million cell updates/sec

Title: US-09-606-910E-1
Perfect score: 113
Sequence: 1 AHGVTSPDTRPAPGSTAPPA 21
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_80:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	113	100.0	347	2 S10571	mucin 1 precursor,
2	113	100.0	1344	1 A35175	mucin 1 precursor,
3	97	85.8	256	2 A60533	tumor-associated a
4	62	54.9	1334	2 T50568	probable multi-dom
5	57	50.4	428	2 T24769	hypothetical prote
6	56	49.6	267	2 F87665	hypothetical prote
7	56	49.6	569	2 A11347	hypothetical prote
8	56	48.6	2706	2 T28155	variant-specific s
9	55	48.7	310	2 A70671	probable puru prot
10	55	48.7	4391	2 A38096	perlecan precursor
11	53	46.9	373	2 A70856	probable lppZ prot
12	53	46.9	417	2 A54416	prostacyclin recep
13	53	46.9	630	2 A39344	tumor-associated m
14	53	46.9	631	2 I52257	episialin - mouse
15	53	46.9	708	2 J43364	gelatinase B (EC 3
16	53	46.9	708	2 S62907	gelatinase B (EC 3
17	52	46.0	115	2 F72570	hypothetical prote
18	52	46.0	182	2 S73046	penicillin binding
19	52	46.0	196	2 T51838	blue copper bindin
20	52	46.0	411	2 D86995	probable D-alanyl-
21	52	46.0	1009	2 J43221	major surface glyc
22	52	46.0	1199	2 J43816	major surface glyc
23	51.5	45.6	635	2 F75477	hypothetical prote
24	51	45.1	217	2 AC1887	hypothetical prote
25	51	45.1	257	2 T10586	small nuclear ribo
26	51	45.1	382	2 T14336	RAD23 protein, iso
27	51	45.1	1201	2 A83007	hypothetical prote
28	50	44.2	402	2 S73773	dihydrolipoamide a
29	50	44.2	418	2 T15142	hypothetical prote

ALIGNMENTS

RESULT 1

S10571
mucin 1 precursor, secreted epithelial tumor antigen splice form - human
N:Contains: mucin 1 secreted breast-cancer-associated splice form
C:Species: Homo sapiens (man)
C:Date: 07-Apr-1994 #sequence revision 07-Apr-1994 #text_change 09-Jul-2004
C:Accession: S10571; JN0100; I56024; S09706; S10217
E:Wreschner, D.H.; Hareuveni, M.; Tsarfaty, I.; Smorodinsky, N.; Horev, J.; Zaretsky, J.
Eur. J. Biochem. 189, 463-473, 1990
A:Title: Human epithelial tumor antigen cDNA sequences. Differential splicing may genera
A:Reference number: S10571; MUID:90276413; PMID:2351132
A:Accession: S10571
A:Molecule type: mRNA
A:Residues: 1-347 <WRE>
A:Cross-references: UNIPROT:Q9UMI8; UNIPARC:UPI0000016B013; EMBL:X52228; NID:G36434; PIDN:
E:Tsarfaty, I.; Hareuveni, M.; Horev, J.; Zaretsky, J.; Weiss, M.; Jeltsch, J.M.; Garnie
Gene 93, 313-318, 1990
A:Title: Isolation and characterization of an expressed hypervariable gene coding for a
A:Reference number: JN0100; MUID:91033045; PMID:1688329
A:Accession: JN0100
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-135; 'Q', 137-142, 'E', 144-163, 204-208, 'A', 210-347 <TSA>
A:Cross-references: UNIPARC:UPI000016A8B8; GB:M35093; NID:G182252; PIDN:AAB59612.1; PID
E:Xing, P.X.; Tjandra, J.J.; Reynolds, K.; McLaughlin, P.J.; Purcell, D.F.J.; McKenzie,
J. Immunol. 142, 3503-3509, 1989
A:Title: Reactivity of anti-human milk fat globule antibodies with synthetic peptides.
A:Reference number: I56024; MUID:89235154; PMID:2715633
A:Accession: I56024
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 182-201 <RES>
A:Cross-references: UNIPARC:UPI000000330DA; GB:M26316; NID:G516622; PIDN:AAA36336.1; PID
E:Tendler, S.J.B.
Biochem. J. 267, 733-737, 1990
A:Title: Elements of secondary structure in a human epithelial mucin core peptide fragme
A:Reference number: S09706; MUID:90253387; PMID:2339983
A:Accession: S09706
A:Molecule type: protein
A:Residues: 182-201 <TEN>
A:Cross-references: UNIPARC:UPI000000330DA
C:Genetics:
A:Gene: GDB:MUC1; PUM
A:Cross-references: GDB:I20705; OMIM:158340
A:Map position: lq21-lq23
C:Keywords: alternative splicing; tandem repeat
E:1-23/Domain: signal sequence #status predicted <SIG>
F:24-347/Product: mucin 1, secreted epithelial tumor antigen splice form #status predict
F:24-163,204-347/Product: mucin 1, secreted breast-cancer-associated splice form #status
Query Match 100.0%; Score 113; DB 2; Length 347;
Best Local Similarity 100.0%; Pred. No. 7.1e-07;

conserved hypothet
hypothetical prote
hypothetical prote
gelatinase B (EC 3
1-phosphatidylinos
1-phosphatidylinos
hypothetical prote
novel cellular pro
hypothetical prote
hypothetical prote
levansucrase (EC 2
cellulosome anchor
pectinesterase hom
serine-repeat anti
probable ATP-depen
diaphanous protein

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AHGVTSPDTRPAPGSTAPPA 21
|||||
Db 175 AHGVTSPDTRPAPGSTAPPA 195

RESULT 2

A35175

mucin 1 precursor, repetitive splice form A [validated] - human

N;Alternate names: breast carcinoma-associated DF3 antigen; core protein KB39; epistatin

N;Alternate names: breast carcinoma-associated DF3 antigen; core protein KB39; epistatin

N;Contains: mucin 1 precursor, epithelial tumor antigen splice form; mucin 1 precursor,

C;Species: Homo sapiens (man)

C;Date: 20-Apr-2000 #sequence revision 20-Apr-2000 #text change 02-Jun-2000

C;Accession: A35175; B35175; A35886; A35887; S10572; S40293; A36735; PX0066; S10218; S51

R;Ligtenberg, M.J.L.; Vos, H.L.; Gennissen, A.M.C.; Hilkens, J.

J. Biol. Chem. 265, 5573-5578, 1990

A;Title: Epistatin, a carcinoma-associated mucin, is generated by a polymorphic gene enc

A;Reference number: A35175; MUID:90202794; PMID:2318825

A;Accession: A35175

A;Molecule type: mRNA

A;Residues: 1-952,1033-1344 <LIG1>

A;Cross-references: UNIPARC:UPI0000174565; GB:M32738; GB:J05288; NID:g182121; PIDN:AAA35

A;Experimental source: splice form A

A;Note: GenBank entries HUMEPISIA1 and HUMEPISIA2 present only the amino-and carboxyl-en

A;Accession: B35175

A;Molecule type: RNA

A;Residues: 1-19,29-952,1033-1344 <LIG2>

A;Cross-references: UNIPARC:UPI0000174566; GB:M32739; GB:J05288; NID:g182126; PIDN:AAA35

A;Experimental source: splice form B

A;Note: GenBank entries HUMEPISIB1 and HUMEPISIB2 present only the amino-and carboxyl-en

R;Gendler, S.J.; Lancaster, C.A.; Taylor-Papadimitriou, J.; Duhig, T.; Peat, N.; Burchel

J. Biol. Chem. 265, 15286-15293, 1990

A;Title: Molecular cloning and expression of human tumor-associated polymorphic epitheli

A;Reference number: A35886; MUID:90368715; PMID:1697589

A;Accession: A35886

A;Status: not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 1-19,29-992,1033-1344 <GEN>

A;Cross-references: UNIPARC:UPI0000174567; GB:J05581; NID:g18869; PIDN:AAA59876.1; PID:

A;Note: GenBank entry HUMMUCHB includes one copy of the tandemly repeated sequence

R;Lan, M.S.; Batra, S.K.; Qi, W.N.; Metzgar, R.S.; Hollingsworth, M.A.

J. Biol. Chem. 265, 15294-15299, 1990

A;Title: Cloning and sequencing of a human pancreatic tumor mucin cDNA.

A;Reference number: A35887; MUID:90368716; PMID:2394722

A;Accession: A35887

A;Status: not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 1-19,29-1109,'S',1111-1339,'A',1341-1344 <LAN>

A;Cross-references: UNIPARC:UPI0000174568; GB:J05582; NID:g189598; PIDN:AAA60019.1; PID:

A;Note: GenBank entry HUMPANMU contains four fewer copies of the tandemly repeated sequ

R;Wreschner, D.H.; Hareuveni, M.; Tsarfaty, I.; Smorodinsky, N.; Horev, J.; Zaretzky, J.

Eur. J. Biochem. 189, 463-473, 1990

A;Title: Human epithelial tumor antigen cDNA sequences. Differential splicing may genera

A;Reference number: S10571; MUID:90276413; PMID:2351132

A;Accession: S10572

A;Molecule type: mRNA

A;Residues: 1-19,29-155,'P',157-175,'P',177-182,'A',184-212,1033-1344 <WRE>

A;Cross-references: UNIPARC:UPI0000174569; EMBL:X52229; NID:g37053

R;Wreschner, D.H.

submitted to the EMBL Data Library, March 1990

A;Reference number: S40293

A;Accession: S40293

A;Molecule type: mRNA

A;Residues: 1-19,29-155,'P',157-175,'P',177-182,'A',184-212,1033-1344 <WR2

A;Cross-references: UNIPARC:UPI000016B0A6; EMBL:X52229; NID:g37053; PIDN:CAA36478.1; PID

R;Abe, M.; Siddiqui, J.; Kufe, D.

Biochem. Biophys. Res. Commun. 165, 644-649, 1989

A;Title: Sequence analysis of the 5' region of the human DF3 breast carcinoma-associated

A;Reference number: A36735; MUID:90088473; PMID:2597151

A;Accession: A36735

A;Molecule type: mRNA

A;Residues: 1-142,'Q',144-162,'Q',164-168 <ABE>

A;Cross-references: UNIPARC:UPI000017456A; EMBL:M31823; NID:g181542; PIDN:AAA35757.1; PI

R;Masuzawa, Y.; Miyauchi, T.; Hamanoue, M.; Ando, S.; Yoshida, J.; Takao, S.; Shimazu, H

J. Biochem. 112, 609-615, 1992

A;Title: A novel core protein as well as polymorphic epithelial mucin carry peanut agglu

A;Reference number: JX0235; MUID:93123189; PMID:1478919

A;Accession: PX0066

A;Molecule type: mRNA

A;Residues: 998-1011,'ES',1014-1017,1018-1032,'T',1034-1037,1038-1057 <MAS>

A;Cross-references: UNIPARC:UPI000017456B; UNIPARC:UPI000017456C; UNIPARC:UPI000017456D

A;Experimental source: gastric carcinoma cell

R;Zrihan-Licht, S.; Baruch, A.; Elroy-Stein, O.; Keydar, I.; Wreschner, D.H.

FEBS Lett. 356, 130-136, 1994

A;Title: Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins cytokine r

A;Reference number: S51026; MUID:95080414; PMID:7988707

A;Contents: annotation

A;Note: undetermined tyrosine residues in the carboxyl-terminal non-repetitive region ar

C;Comment: This protein is length polymorphic. Individuals may have between 21 and 125 c

partial repeats. The repeat shown is defined by SmaI nuclease sites.

C;Comment: Serine and threonine residues in the tandem repeat domain are extensively gly

C;Comment: For an alternative splice form without a tandem repeat domain, see PIR:S48146

C;Genetics:

A;Gene: GDB:MUC1; PUM

A;Cross-references: GDB:I20705; OMIM:158340

A;Map position: 1q21-1q23

A;Intons: 20/1; 82/3; 1165/3; 1184/2; 1230/1; 1270/3; 1320/3

C;Superfamily: polymorphic epithelial mucin

C;Keywords: alternative splicing; mucin 1 precursor, splice form A #status predicted <PREA>

F;1-1344/Product: mucin 1 precursor, signal sequence #link PREA #status predicted <SIGA>

F;1-62/Region: mucin 1 amino-terminal non-repetitive

F;1-23/Domin: signal sequence #link PREA #status predicted <SIGA>

F;1-19,29-32/Domin: signal sequence #link PREA #status predicted <SIGB>

F;1-19,29-1344/Product: mucin 1 precursor, splice form B #status predicted <PREB>

F;1-19,29-212,1033-1344/Product: mucin 1 precursor, epithelial tumor antigen splice form

F;138-1017/Region: 20-residue repeats (GSTAPPAGVTSAPDTRPPAP)

F;1143-1344/Region: mucin 1 carboxyl-terminal non-repetitive

F;1245-1272/Domin: transmembrane #status predicted <TRM>

F;1046,1064,1118,1144,1222/Binding site: carboxydrate (Asn) (covalent) #status predicted

F;1213/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 100.0%; Score 113; DB 1; Length 1344;

Best Local Similarity 100.0%; Pred. No. 2.5e-06;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AHGVTSPDTRPAPGSTAPPA 21
|||||
Db 144 AHGVTSPDTRPAPGSTAPPA 164

RESULT 3

A60533

tumor-associated antigen DF3 - human

C;Species: Homo sapiens (man)

C;Date: 19-Mar-1993 #sequence revision 07-May-1993 #text change 05-Oct-2004

C;Accession: A60533

R;Merlo, G.R.; Siddiqui, J.; Cropp, C.S.; Liscia, D.S.; Lidereau, R.; Callahan, R.; Kufe

Cancer Res. 49, 6966-6971, 1989

A;Title: Frequent alteration of the DF3 tumor-associated antigen gene in primary human t

A;Reference number: A60533; MUID:90058554; PMID:2582438

A;Accession: A60533

A;Status: not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 1-256 <MER>

A;Cross-references: UNIPROT:Q7M4M7; UNIPARC:UPI0000177CC7

C;Genetics:

A;Map position: 1q21-q24

C;Keywords: glycoprotein; tandem repeat

Query Match 85.8%; Score 97; DB 2; Length 256;

Best Local Similarity 85.7%; Pred. No. 5.1e-05;

Matches 18; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AHGVTSPDTRPAPGSTAPPA 21

F87565
hypothetical protein CC3360 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #tax_change 09-Jul-2004
C:Accession: F87565
R:Nierman, W.C.; DeBlyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.C.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of *Caulobacter crescentus*.
A:Reference number: A87249; MUID:21173698; PMID:11259647

Qy 7 APDTRPAPGSTAPPA 21
|||||I
Db 2449 APDTRPVSPPLPPA 2463

RESULT 9
A70671
probable purU protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 05-Oct-2004
C:Accession: A70671; S73058
R:Colle, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: A70671
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-310 <COL>
A:Cross-references: UNIPROT:Q50453; UNIPARC:UPI0000132B74; GB:AL123456; NID:9118-309/Domain: phosphoribosylglycinamide formyltransferase homology <PRGF>
R:Smith, D.R.; Robison, K.
submitted to the EMBL Data Library, September 1994
A:Description: Mycobacterium tuberculosis cosmid tbc2.
A:Reference number: S73053
A:Accession: S73058
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-310 <SMI>
A:Cross-references: UNIPARC:UPI0000132B74; EMBL:U00024; NID:G560506; PIDN:AAA50945.1; PIDN:AAA50945.1; PIDN:AAA50945.1
C:Genetics:
A:Gene: purU
C:Superfamily: formyltetrahydrofolate deformylase; phosphoribosylglycinamide formyltransferase
F:118-309/Domain: phosphoribosylglycinamide formyltransferase homology <PRGF>

Query Match 48.7%; Score 55; DB 2; Length 310;
Best Local Similarity 47.6%; Pred. No. 9, 7;
Matches 10; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

Qy 1 AHGVTSPDTRPAPGSTAPPA 21
|||:|||||
Db 8 AHATNEPDYPPGGPPPA 28

RESULT 10
A38096
N:perlecan precursor - human
N:Alternate names: basement membrane heparan sulfate proteoglycan; heparan sulfate proteoglycan
C:Species: Homo sapiens (man)
C:Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 12-Jul-2004
C:Accession: A38096; S19256; S77946; A41059; A40306; B33625; A33625; A41736
R:Murdoch, A.D.; Dodge, G.R.; Cohen, I.; Tuan, R.S.; Iozzo, R.V. J. Biol. Chem. 267, 8544-8557, 1992
A:Title: Primary structure of the human heparan sulfate proteoglycan from basement membrane, laminin, neural cell adhesion molecules, and epidermal growth factor.
A:Reference number: A38096; MUID:92235084; PMID:1569102
A:Accession: A38096
A:Molecule type: mRNA
A:Residues: 1-4391 <MUR>
A:Cross-references: UNIPROT:P98160; UNIPARC:UPI0000168756; GB:M85289; NID:G184426; PIDN:G184426; PIDN:G184426
R:Kallunki, P.; Tryggvason, K.
J. Cell Biol. 116, 559-571, 1992
A:Title: Human basement membrane heparan sulfate proteoglycan core protein: a 467-kD protein with multiple cell adhesion molecules, and epidermal growth factor.
A:Reference number: A41736; MUID:92112994; PMID:1730768
A:Accession: S19256
A:Molecule type: mRNA
A:Residues: 1-57, 'D', '59-434, 'A', '436, 'FL', '438-449, 'Q', '451-502, 'A', '503-792, 'K', '794-908, 'R', '71-2979, 'H', '2981-2994, 'G', '2996-3167, 'T', '3169-3240, 'R', '3242-3426, 'R', '3428-3631, 'Q', '3633-4373, 'P', '4374-4391, 'G', '4392-4400, 'R', '4401-4410, 'K', '4411-4420, 'R', '4421-4430, 'K', '4431-4440, 'R', '4441-4450, 'K', '4451-4460, 'R', '4461-4470, 'K', '4471-4480, 'R', '4481-4490, 'K', '4491-4500, 'R', '4501-4510, 'K', '4511-4520, 'R', '4521-4530, 'K', '4531-4540, 'R', '4541-4550, 'K', '4551-4560, 'R', '4561-4570, 'K', '4571-4580, 'R', '4581-4590, 'K', '4591-4600, 'R', '4601-4610, 'K', '4611-4620, 'R', '4621-4630, 'K', '4631-4640, 'R', '4641-4650, 'K', '4651-4660, 'R', '4661-4670, 'K', '4671-4680, 'R', '4681-4690, 'K', '4691-4700, 'R', '4701-4710, 'K', '4711-4720, 'R', '4721-4730, 'K', '4731-4740, 'R', '4741-4750, 'K', '4751-4760, 'R', '4761-4770, 'K', '4771-4780, 'R', '4781-4790, 'K', '4791-4800, 'R', '4801-4810, 'K', '4811-4820, 'R', '4821-4830, 'K', '4831-4840, 'R', '4841-4850, 'K', '4851-4860, 'R', '4861-4870, 'K', '4871-4880, 'R', '4881-4890, 'K', '4891-4900, 'R', '4901-4910, 'K', '4911-4920, 'R', '4921-4930, 'K', '4931-4940, 'R', '4941-4950, 'K', '4951-4960, 'R', '4961-4970, 'K', '4971-4980, 'R', '4981-4990, 'K', '4991-5000, 'R', '5001-5010, 'K', '5011-5020, 'R', '5021-5030, 'K', '5031-5040, 'R', '5041-5050, 'K', '5051-5060, 'R', '5061-5070, 'K', '5071-5080, 'R', '5081-5090, 'K', '5091-5100, 'R', '5101-5110, 'K', '5111-5120, 'R', '5121-5130, 'K', '5131-5140, 'R', '5141-5150, 'K', '5151-5160, 'R', '5161-5170, 'K', '5171-5180, 'R', '5181-5190, 'K', '5191-5200, 'R', '5201-5210, 'K', '5211-5220, 'R', '5221-5230, 'K', '5231-5240, 'R', '5241-5250, 'K', '5251-5260, 'R', '5261-5270, 'K', '5271-5280, 'R', '5281-5290, 'K', '5291-5300, 'R', '5301-5310, 'K', '5311-5320, 'R', '5321-5330, 'K', '5331-5340, 'R', '5341-5350, 'K', '5351-5360, 'R', '5361-5370, 'K', '5371-5380, 'R', '5381-5390, 'K', '5391-5400, 'R', '5401-5410, 'K', '5411-5420, 'R', '5421-5430, 'K', '5431-5440, 'R', '5441-5450, 'K', '5451-5460, 'R', '5461-5470, 'K', '5471-5480, 'R', '5481-5490, 'K', '5491-5500, 'R', '5501-5510, 'K', '5511-5520, 'R', '5521-5530, 'K', '5531-5540, 'R', '5541-5550, 'K', '5551-5560, 'R', '5561-5570, 'K', '5571-5580, 'R', '5581-5590, 'K', '5591-5600, 'R', '5601-5610, 'K', '5611-5620, 'R', '5621-5630, 'K', '5631-5640, 'R', '5641-5650, 'K', '5651-5660, 'R', '5661-5670, 'K', '5671-5680, 'R', '5681-5690, 'K', '5691-5700, 'R', '5701-5710, 'K', '5711-5720, 'R', '5721-5730, 'K', '5731-5740, 'R', '5741-5750, 'K', '5751-5760, 'R', '5761-5770, 'K', '5771-5780, 'R', '5781-5790, 'K', '5791-5800, 'R', '5801-5810, 'K', '5811-5820, 'R', '5821-5830, 'K', '5831-5840, 'R', '5841-5850, 'K', '5851-5860, 'R', '5861-5870, 'K', '5871-5880, 'R', '5881-5890, 'K', '5891-5900, 'R', '5901-5910, 'K', '5911-5920, 'R', '5921-5930, 'K', '5931-5940, 'R', '5941-5950, 'K', '5951-5960, 'R', '5961-5970, 'K', '5971-5980, 'R', '5981-5990, 'K', '5991-6000, 'R', '6001-6010, 'K', '6011-6020, 'R', '6021-6030, 'K', '6031-6040, 'R', '6041-6050, 'K', '6051-6060, 'R', '6061-6070, 'K', '6071-6080, 'R', '6081-6090, 'K', '6091-6100, 'R', '6101-6110, 'K', '6111-6120, 'R', '6121-6130, 'K', '6131-6140, 'R', '6141-6150, 'K', '6151-6160, 'R', '6161-6170, 'K', '6171-6180, 'R', '6181-6190, 'K', '6191-6200, 'R', '6201-6210, 'K', '6211-6220, 'R', '6221-6230, 'K', '6231-6240, 'R', '6241-6250, 'K', '6251-6260, 'R', '6261-6270, 'K', '6271-6280, 'R', '6281-6290, 'K', '6291-6300, 'R', '6301-6310, 'K', '6311-6320, 'R', '6321-6330, 'K', '6331-6340, 'R', '6341-6350, 'K', '6351-6360, 'R', '6361-6370, 'K', '6371-6380, 'R', '6381-6390, 'K', '6391-6400, 'R', '6401-6410, 'K', '6411-6420, 'R', '6421-6430, 'K', '6431-6440, 'R', '6441-6450, 'K', '6451-6460, 'R', '6461-6470, 'K', '6471-6480, 'R', '6481-6490, 'K', '6491-6500, 'R', '6501-6510, 'K', '6511-6520, 'R', '6521-6530, 'K', '6531-6540, 'R', '6541-6550, 'K', '6551-6560, 'R', '6561-6570, 'K', '6571-6580, 'R', '6581-6590, 'K', '6591-6600, 'R', '6601-6610, 'K', '6611-6620, 'R', '6621-6630, 'K', '6631-6640, 'R', '6641-6650, 'K', '6651-6660, 'R', '6661-6670, 'K', '6671-6680, 'R', '6681-6690, 'K', '6691-6700, 'R', '6701-6710, 'K', '6711-6720, 'R', '6721-6730, 'K', '6731-6740, 'R', '6741-6750, 'K', '6751-6760, 'R', '6761-6770, 'K', '6771-6780, 'R', '6781-6790, 'K', '6791-6800, 'R', '6801-6810, 'K', '6811-6820, 'R', '6821-6830, 'K', '6831-6840, 'R', '6841-6850, 'K', '6851-6860, 'R', '6861-6870, 'K', '6871-6880, 'R', '6881-6890, 'K', '6891-6900, 'R', '6901-6910, 'K', '6911-6920, 'R', '6921-6930, 'K', '6931-6940, 'R', '6941-6950, 'K', '6951-6960, 'R', '6961-6970, 'K', '6971-6980, 'R', '6981-6990, 'K', '6991-7000, 'R', '7001-7010, 'K', '7011-7020, 'R', '7021-7030, 'K', '7031-7040, 'R', '7041-7050, 'K', '7051-7060, 'R', '7061-7070, 'K', '7071-7080, 'R', '7081-7090, 'K', '7091-7100, 'R', '7101-7110, 'K', '7111-7120, 'R', '7121-7130, 'K', '7131-7140, 'R', '7141-7150, 'K', '7151-7160, 'R', '7161-7170, 'K', '7171-7180, 'R', '7181-7190, 'K', '7191-7200, 'R', '7201-7210, 'K', '7211-7220, 'R', '7221-7230, 'K', '7231-7240, 'R', '7241-7250, 'K', '7251-7260, 'R', '7261-7270, 'K', '7271-7280, 'R', '7281-7290, 'K', '7291-7300, 'R', '7301-7310, 'K', '7311-7320, 'R', '7321-7330, 'K', '7331-7340, 'R', '7341-7350, 'K', '7351-7360, 'R', '7361-7370, 'K', '7371-7380, 'R', '7381-7390, 'K', '7391-7400, 'R', '7401-7410, 'K', '7411-7420, 'R', '7421-7430, 'K', '7431-7440, 'R', '7441-7450, 'K', '7451-7460, 'R', '7461-7470, 'K', '7471-7480, 'R', '7481-7490, 'K', '7491-7500, 'R', '7501-7510, 'K', '7511-7520, 'R', '7521-7530, 'K', '7531-7540, 'R', '7541-7550, 'K', '7551-7560, 'R', '7561-7570, 'K', '7571-7580, 'R', '7581-7590, 'K', '7591-7600, 'R', '7601-7610, 'K', '7611-7620, 'R', '7621-7630, 'K', '7631-7640, 'R', '7641-7650, 'K', '7651-7660, 'R', '7661-7670, 'K', '7671-7680, 'R', '7681-7690, 'K', '7691-7700, 'R', '7701-7710, 'K', '7711-7720, 'R', '7721-7730, 'K', '7731-7740, 'R', '7741-7750, 'K', '7751-7760, 'R', '7761-7770, 'K', '7771-7780, 'R', '7781-7790, 'K', '7791-7800, 'R', '7801-7810, 'K', '7811-7820, 'R', '7821-7830, 'K', '7831-7840, 'R', '7841-7850, 'K', '7851-7860, 'R', '7861-7870, 'K', '7871-7880, 'R', '7881-7890, 'K', '7891-7900, 'R', '7901-7910, 'K', '7911-7920, 'R', '7921-7930, 'K', '7931-7940, 'R', '7941-7950, 'K', '7951-7960, 'R', '7961-7970, 'K', '7971-7980, 'R', '7981-7990, 'K', '7991-8000, 'R', '8001-8010, 'K', '8011-8020, 'R', '8021-8030, 'K', '8031-8040, 'R', '8041-8050, 'K', '8051-8060, 'R', '8061-8070, 'K', '8071-8080, 'R', '8081-8090, 'K', '8091-8100, 'R', '8101-8110, 'K', '8111-8120, 'R', '8121-8130, 'K', '8131-8140, 'R', '8141-8150, 'K', '8151-8160, 'R', '8161-8170, 'K', '8171-8180, 'R', '8181-8190, 'K', '8191-8200, 'R', '8201-8210, 'K', '8211-8220, 'R', '8221-8230, 'K', '8231-8240, 'R', '8241-8250, 'K', '8251-8260, 'R', '8261-8270, 'K', '8271-8280, 'R', '8281-8290, 'K', '8291-8300, 'R', '8301-8310, 'K', '8311-8320, 'R', '8321-8330, 'K', '8331-8340, 'R', '8341-8350, 'K', '8351-8360, 'R', '8361-8370, 'K', '8371-8380, 'R', '8381-8390, 'K', '8391-8400, 'R', '8401-8410, 'K', '8411-8420, 'R', '8421-8430, 'K', '8431-8440, 'R', '8441-8450, 'K', '8451-8460, 'R', '8461-8470, 'K', '8471-8480, 'R', '8481-8490, 'K', '8491-8500, 'R', '8501-8510, 'K', '8511-8520, 'R', '8521-8530, 'K', '8531-8540, 'R', '8541-8550, 'K', '8551-8560, 'R', '8561-8570, 'K', '8571-8580, 'R', '8581-8590, 'K', '8591-8600, 'R', '8601-8610, 'K', '8611-8620, 'R', '8621-8630, 'K', '8631-8640, 'R', '8641-8650, 'K', '8651-8660, 'R', '8661-8670, 'K', '8671-8680, 'R', '8681-8690, 'K', '8691-8700, 'R', '8701-8710, 'K', '8711-8720, 'R', '8721-8730, 'K', '8731-8740, 'R', '8741-8750, 'K', '8751-8760, 'R', '8761-8770, 'K', '8771-8780, 'R', '8781-8790, 'K', '8791-8800, 'R', '8801-8810, 'K', '8811-8820, 'R', '8821-8830, 'K', '8831-8840, 'R', '8841-8850, 'K', '8851-8860, 'R', '8861-8870, 'K', '8871-8880, 'R', '8881-8890, 'K', '8891-8900, 'R', '8901-8910, 'K', '8911-8920, 'R', '8921-8930, 'K', '8931-8940, 'R', '8941-8950, 'K', '8951-8960, 'R', '8961-8970, 'K', '8971-8980, 'R', '8981-8990, 'K', '8991-9000, 'R', '9001-9010, 'K', '9011-9020, 'R', '9021-9030, 'K', '9031-9040, 'R', '9041-9050, 'K', '9051-9060, 'R', '9061-9070, 'K', '9071-9080, 'R', '9081-9090, 'K', '9091-9100, 'R', '9101-9110, 'K', '9111-9120, 'R', '9121-9130, 'K', '9131-9140, 'R', '9141-9150, 'K', '9151-9160, 'R', '9161-9170, 'K', '9171-9180, 'R', '9181-9190, 'K', '9191-9200, 'R', '9201-9210, 'K', '9211-9220, 'R', '9221-9230, 'K', '9231-9240, 'R', '9241-9250, 'K', '9251-9260, 'R', '9261-9270, 'K', '9271-9280, 'R', '9281-9290, 'K', '9291-9300, 'R', '9301-9310, 'K', '9311-9320, 'R', '9321-9330, 'K', '9331-9340, 'R', '9341-9350, 'K', '9351-9360, 'R', '9361-9370, 'K', '9371-9380, 'R', '9381-9390, 'K', '9391-9400, 'R', '9401-9410, 'K', '9411-9420, 'R', '9421-9430, 'K', '9431-9440, 'R', '9441-9450, 'K', '9451-9460, 'R', '9461-9470, 'K', '9471-9480, 'R', '9481-9490, 'K', '9491-9500, 'R', '9501-9510, 'K', '9511-9520, 'R', '9521-9530, 'K', '9531-9540, 'R', '9541-9550, 'K', '9551-9560, 'R', '9561-9570, 'K', '9571-9580, 'R', '9581-9590, 'K', '9591-9600, 'R', '9601-9610, 'K', '9611-9620, 'R', '9621-9630, 'K', '9631-9640, 'R', '9641-9650, 'K', '9651-9660, 'R', '9661-9670, 'K', '9671-9680, 'R', '9681-9690, 'K', '9691-9700, 'R', '9701-9710, 'K', '9711-9720, 'R', '9721-9730, 'K', '9731-9740, 'R', '9741-9750, 'K', '9751-9760, 'R', '9761-9770, 'K', '9771-9780, 'R', '9781-9790, 'K', '9791-9800, 'R', '9801-9810, 'K', '9811-9820, 'R', '9821-9830, 'K', '9831-9840, 'R', '9841-9850, 'K', '9851-9860, 'R', '9861-9870, 'K', '9871-9880, 'R', '9881-9890, 'K', '9891-9900, 'R', '9901-9910, 'K', '9911-9920, 'R', '9921-9930, 'K', '9931-9940, 'R', '9941-9950, 'K', '9951-9960, 'R', '9961-9970, 'K', '9971-9980, 'R', '9981-9990, 'K', '9991-10000, 'R', '10001-10010, 'K', '10011-10020, 'R', '10021-10030, 'K', '10031-10040, 'R', '10041-10050, 'K', '10051-10060, 'R', '10061-10070, 'K', '10071-10080, 'R', '10081-10090, 'K', '10091-10100, 'R', '10101-10110, 'K', '10111-10120, 'R', '10121-10130, 'K', '10131-10140, 'R', '10141-10150, 'K', '10151-10160, 'R', '10161-10170, 'K', '10171-10180, 'R', '10181-10190, 'K', '10191-10200, 'R', '10201-10210, 'K', '10211-10220, 'R', '10221-10230, 'K', '10231-10240, 'R', '10241-10250, 'K', '10251-10260, 'R', '10261-10270, 'K', '10271-10280, 'R', '10281-10290, 'K', '10291-10300, 'R', '10301-10310, 'K', '10311-10320, 'R', '10321-10330, 'K', '10331-10340, 'R', '10341-10350, 'K', '10351-10360, 'R', '10361-10370, 'K', '10371-10380, 'R', '10381-10390, 'K', '10391-10400, 'R', '10401-10410, 'K', '10411-10420, 'R', '10421-10430, 'K', '10431-10440, 'R', '10441-10450, 'K', '10451-10460, 'R', '10461-10470, 'K', '10471-10480, 'R', '10481-10490, 'K', '10491-10500, 'R', '10501-10510, 'K', '10511-10520, 'R', '10521-10530, 'K', '10531-10540, 'R', '10541-10550, 'K', '10551-10560, 'R', '10561-10570, 'K', '10571-10580, 'R', '10581-10590, 'K', '10591-10600, 'R', '10601-10610, 'K', '10611-10620, 'R', '10621-10630, 'K', '10631-10640, 'R', '10641-10650, 'K', '10651-10660, 'R', '10661-10670, 'K', '10671-10680, 'R', '10681-10690, 'K', '10691-10700, 'R', '10701-10710, 'K', '10711-10720, 'R', '10721-10730, 'K', '10731-10740, 'R', '10741-10750, 'K', '10751-10760, 'R', '10761-10770, 'K', '10771-10780, 'R', '10781-10790, 'K', '10791-10800, 'R', '10801-10810, 'K', '10811-10820, 'R', '10821-10830, 'K', '10831-10840, 'R', '10841-10850, 'K', '10851-10860, 'R', '10861-10870, 'K', '10871-10880, 'R', '10881-10890, 'K', '10891-10900, 'R', '10901-10910, 'K', '10911-10920, 'R', '10921-10930, 'K', '10931-10940, 'R', '10941-10950, 'K', '10951-10960, 'R', '10961-10970, 'K', '10971-10980, 'R', '10981-10990, 'K', '10991-11000, 'R', '11001-11010, 'K', '11011-11020, 'R', '11021-11030, 'K', '11031-11040, 'R', '11041-11050, 'K', '11051-11060, 'R', '11061-11070, 'K', '11071-11080, 'R', '11081-11090, 'K', '11091-11100, 'R', '11101-11110, 'K', '11111-11120, 'R', '11121-11130, 'K', '11131-11140, 'R', '11141-11150, 'K', '11151-11160, 'R', '11161-11170, 'K', '11171-11180, 'R', '11181-11190, 'K', '11191-11200, 'R', '11201-11210, 'K', '11211-11220, 'R', '11221-11230, 'K', '11231-11240, 'R', '11241-11250, 'K', '11251-11260, 'R', '11261-11270, 'K', '11271-11280, 'R', '11281-11290, 'K', '11291-11300, 'R', '11301-11310, 'K', '11311-11320, 'R', '11321-11330, 'K', '11331-11340, 'R', '11341-11350, 'K', '11351-11360, 'R', '11361-11370, 'K', '11371-11380, 'R', '11381-11390, 'K', '11391-11400, 'R', '11401-11410, 'K', '11411-11420, 'R', '11421-11430, 'K', '11431-11440, 'R', '11441-11450, 'K', '11451-11460, 'R', '11461-11470, 'K', '11471-11480, 'R', '11481-11490, 'K', '11491-11500, 'R', '11501-11510, 'K', '11511-11520, 'R', '11521-11530, 'K', '11531-11540, 'R', '11541-11550, 'K', '11551-11560, 'R', '11561-11570, 'K', '11571-11580, 'R', '11581-11590, 'K', '11591-11600, 'R', '11601-11610, 'K', '11611-11620, 'R', '11621-11630, 'K', '11631-11640, 'R', '11641-11650, 'K', '11651-11660, 'R', '11661-11670, 'K', '11671-11680, 'R', '11681-11690, 'K', '11691-11700, 'R', '117

QY 2 HGVTSAPDTRPAPGSTAP 19
 ||| ||| ||| ||| |||
 Db 2136 HCHTSGSPSTPVPGSTP 2153
 ||| ||| ||| ||| |||

RESULT 11
 A70856
 probable lppZ protein - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Jul-2004
 C:Accession: A70856
 R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
 A: Authors: Sgates, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A: Reference number: A70500; MUID: 98295987; PMID: 9634230
 A: Accession: A70856
 A: Status: preliminary; nucleic acid sequence not shown; translation not shown
 A: Molecule type: DNA
 A: Residues: 1-373 <COL>
 A: Cross-references: UNIPROT: O53253; UNIPARC: UPI00000318FB; GB: AL021287; GB: AL123456; NID:
 A: Experimental source: strain H37RV
 C: Geneticks:
 A: Gene: lppZ

Query Match 46.9%; Score 53; DB 2; Length 373;
 Best Local Similarity 56.2%; Pred. No. 20;
 Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 5 TSAPDTRPAPGSTAPP 20
 ||| ||| ||| ||| |||
 Db 36 TTEPRLQPSTSTPP 51
 ||| ||| ||| ||| |||

RESULT 12
 A54416
 prostacyclin receptor - mouse
 C:Species: Mus musculus (house mouse)
 C>Date: 09-Sep-1994 #sequence_revision 09-Sep-1994 #text_change 20-Jun-2000
 C:Accession: A54416
 R: Namba, T.; Oida, H.; Sugimoto, Y.; Kakizuka, A.; Negishi, M.; Ichikawa, A.; Narumiya, J. Biol. Chem. 269, 9986-9992, 1994
 A: Title: cDNA cloning of a mouse prostacyclin receptor. Multiple signaling pathways and
 A: Reference number: A54416; MUID: 94193694; PMID: 7511597
 A: Accession: A54416
 A: Status: preliminary
 A: Molecule type: mRNA
 A: Residues: 1-417 <NAM>
 A: Cross-references: UNIPARC: UPI000016CFCE; GB: D26157; NID: 9493687; PIDN: BAA05144.1; PID:
 C: Superfamily: prostaglandin E receptor EPI

Query Match 46.9%; Score 53; DB 2; Length 417;
 Best Local Similarity 55.0%; Pred. No. 23;
 Matches 11; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 2 HGVTSAPDTRPAPGSTAP 21
 ||| ||| ||| ||| |||
 Db 345 HGDLOALPSRPSGRDRPPA 364
 ||| ||| ||| ||| |||

RESULT 13
 A39344
 tumor-associated mucin (MUC1) homolog precursor - mouse
 C:Species: Mus musculus (house mouse)
 C>Date: 03-Apr-1992 #sequence_revision 03-Apr-1992 #text_change 09-Jul-2004
 C:Accession: A39344
 R: Spicer, A.P.; Parry, G.; Patton, S.; Gendler, S.J.
 J. Biol. Chem. 266, 15099-15109, 1991.
 A: Title: Molecular cloning and analysis of the mouse homologue of the tumor-associated m
 a loss of minisatellite-like polymorphism.
 A: Reference number: A39344; MUID: 91332029; PMID: 1714452

A: Accession: A39344
 A: Status: preliminary
 A: Molecule type: DNA
 A: Residues: 1-630 <SPI>
 A: Cross-references: UNIPROT: Q02496; UNIPARC: UPI0000029422; GB: M64928
 C: Keywords: cytoskeleton; transmembrane protein

Query Match 46.9%; Score 53; DB 2; Length 630;
 Best Local Similarity 55.0%; Pred. No. 33;
 Matches 11; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 1 AHGVTSPDTRPAPGSTAPP 20
 ||| ||| ||| ||| |||
 Db 233 AHGGTSPATSPLRDSTSSP 252
 ||| ||| ||| ||| |||

RESULT 14
 I52257
 episialin - mouse
 C:Species: Mus musculus (house mouse)
 C>Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 09-Jul-2004
 C:Accession: I52257; I65210
 R: Vos, H.L.; De Vries, Y.; Hilkens, J.
 Biochem. Biophys. Res. Commun. 181, 121-130, 1991
 A: Title: The mouse episialin (Muc1) gene and its promoter. Rapid evolution of the repeat
 A: Reference number: I52257; MUID: 92068178; PMID: 1958179
 A: Accession: I52257
 A: Status: preliminary; translated from GB/EMBL/DBDJ
 A: Molecule type: DNA
 A: Residues: 1-631 <RES>
 A: Cross-references: UNIPROT: Q02496; UNIPARC: UPI000002A11B; GB: M77226; NID: g199835; PIDN:
 A: Accession: I65210
 A: Status: preliminary; translated from GB/EMBL/DBDJ
 A: Molecule type: mRNA
 A: Residues: 1-631 <RE2>
 A: Cross-references: UNIPARC: UPI000002A11B; GB: M84683; NID: g199842; PIDN: AAA39756.1; PID:
 C: Geneticks:
 A: Gene: Muc1
 A: Introns: 20/1; 454/3; 472/2; 517/1; 557/3; 607/3

Query Match 46.9%; Score 53; DB 2; Length 631;
 Best Local Similarity 55.0%; Pred. No. 33;
 Matches 11; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 1 AHGVTSPDTRPAPGSTAPP 20
 ||| ||| ||| ||| |||
 Db 234 AHGGTSPATSPLRDSTSSP 253
 ||| ||| ||| ||| |||

RESULT 15
 JC4364
 gelatinase B (EC 3.4.24.35) precursor - rat
 N: Alternate names: 92-kDa type IV collagenase; matrix metalloproteinase 9 (MMP9)
 C:Species: Rattus norvegicus (Norway rat)
 C>Date: 10-Jan-1996 #sequence_revision 08-Feb-1996 #text_change 09-Jul-2004
 C:Accession: JC4364
 R: Okada, A.; Santavicca, M.; Basset, P.
 Gene 164, 317-321, 1995
 A: Title: The cDNA cloning and expression of the gene encoding rat gelatinase B.
 A: Reference number: JC4364; MUID: 96069602; PMID: 7590350
 A: Accession: JC4364
 A: Molecule type: mRNA
 A: Residues: 1-708 <OKA>
 A: Cross-references: UNIPROT: P50282; UNIPARC: UPI00001679CE; GB: U24441; NID: g1173505; PID:
 A: Experimental source: skin wounds
 C: Comment: This enzyme belongs to the matrix metalloproteinase family, proteolytic enzy
 ogression.
 C: Geneticks:
 A: Gene: GelB
 C: Superfamily: gelatinase A; fibronectin type II repeat homology; hemopexin repeat homo
 C: Keywords: extracellular matrix; fibroblast; glycoprotein; hydrolase; metalloproteinase;
 F1-24/Domain: signal sequence #status predicted <SIG>
 F125-708/Product: progelatinase B #status predicted <PRO>

Tue Apr 18 08:15:24 2006

F:25-107/Domain: activation peptide #status predicted <ACT>
F:108-708/Product: gelatinase B #status predicted <MAT>
F:214-389/Region: collagen binding #status predicted
F:231-272/Domain: fibronectin type II repeat homology <2F1>
F:289-330/Domain: fibronectin type II repeat homology <2F2>
F:348-389/Domain: fibronectin type II repeat homology <2F3>
F:514-707/Domain: hemopexin repeat homology <PXW>
F:39,121/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:100,402,406,412/Binding site: zinc, catalytic (Cys, His, His) (inhibited) #status
F:402,406,412/Binding site: zinc, catalytic (His) (active) #status predicted
F:403/Active site: Glu #status predicted
F:519-707/Disulfide bonds: #status predicted

Query Match 46.9%; Score 53; DB 2; Length 708;
Best Local Similarity 55.6%; Pred. No. 37;
Matches 10; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 3 GVTAPDTRPAPGSTAPP 20
| | | | | | | | | |
| | | | | | | | | |
Db 483 GPTVAPTGAPSGPTGPP 500

Search completed: April 14, 2006, 10:30:25
Job time : 26.5854 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2006, 10:13:30 ; Search time 311.927 Seconds
(without alignments)
29.581 Million cell updates/sec

Title: US-09-606-910E-1
Perfect score: 113
Sequence: 1 AHGVTAPDTRPAGSTAPPA 21

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A. Genesep 21.*

- 1: genesep1980s.*
- 2: genesep1990s.*
- 3: genesep2000s.*
- 4: genesep2001s.*
- 5: genesep2002s.*
- 6: genesep2003as.*
- 7: genesep2003bs.*
- 8: genesep2004s.*
- 9: genesep2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	113	100.0	21	2	AAV25111
2	113	100.0	21	3	AAV96171
3	113	100.0	21	3	AAV96173
4	113	100.0	21	3	AAV96169
5	113	100.0	21	3	AAV96180
6	113	100.0	21	3	AAV96179
7	113	100.0	21	3	AAV96172
8	113	100.0	21	3	AAV96174
9	113	100.0	21	3	AAV96176
10	113	100.0	21	3	AAV96178
11	113	100.0	21	3	AAV96177
12	113	100.0	21	3	AAV96175
13	113	100.0	21	8	ADK65929
14	113	100.0	21	8	ADK65929
15	113	100.0	25	3	ADC78317
16	113	100.0	25	6	ABU07528
17	113	100.0	25	7	ADK25858
18	113	100.0	25	8	ADK32636
19	113	100.0	25	8	ADK58858
20	113	100.0	28	2	AAW03361
21	113	100.0	30	8	ADK65927
22	113	100.0	31	4	AAW46087
23	113	100.0	40	2	AAW68002
24	113	100.0	40	2	AAW54873

25	113	100.0	40	2	AAW72703
26	113	100.0	40	5	ABP56039
27	113	100.0	40	8	ADP32639
28	113	100.0	40	9	ADK05275
29	113	100.0	40	9	ADY53503
30	113	100.0	41	3	AAV96170
31	113	100.0	43	6	AAE33952
32	113	100.0	43	7	ADD88870
33	113	100.0	46	5	ABP56035
34	113	100.0	46	6	AAE33936
35	113	100.0	46	9	ADY53499
36	113	100.0	50	2	AAW35739
37	113	100.0	51	2	AAW31697
38	113	100.0	100	5	ABB76181
39	113	100.0	105	2	AAE68022
40	113	100.0	105	2	AAW72697
41	113	100.0	109	8	ADI57744
42	113	100.0	159	9	ADK05280
43	113	100.0	173	3	AAV71021
44	113	100.0	216	3	AAV92665
45	113	100.0	295	3	AAV71027

ALIGNMENTS

RESULT 1

AAV25111
ID AAV25111 standard; peptide; 21 AA.
XX
AC AAV25111;
XX
DT 25-AUG-1999 (first entry)
XX
DE Synthetic glycopeptide #1 homologous to MUC1.
XX
KW Immunodominant; epitope; MUC1; epithelial mucin; tumour; vaccine;
KW mammary carcinoma; colorectal carcinoma; pancreas carcinoma.
XX
OS Synthetic.
XX
PN DE19758400-A1.
XX
PD 01-JUL-1999.
XX
PF 30-DEC-1997; 97DE-01058400.
XX
PR 30-DEC-1997; 97DE-01058400.
XX
PA (DELB-) DELBRUECK CENT MOLEKULARE MEDIZIN MAX.
(HANI/) HANISCH F.
XX
PI Karsten U, Hanisch F, Paulsen H;
XX
DR WPI; 1999-372374/32.
XX
PT New tumour vaccine - for treating tumour cells of mammary tumours,
colorectal and pancreas carcinomas.
XX
PS Example 1; Page 3; 6pp; German.
XX
CC This invention describes a novel tumour vaccine which comprises a
synthetic peptide derived from human epithelial mucin MUC1. This active
tumour vaccine can be used against tumour cells from mammary, colorectal
or pancreas carcinomas. This sequence represents a synthetic glycopeptide
CC which is homologous to an immunodominant epitope from human MUC1
XX
SQ Sequence 21 AA;

Query Match 100.0%; Score 113; DB 2; Length 21;
Best Local Similarity 100.0%; Pred. No. 5e-07;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY      1 AHGVTSPDTRPAGSTAPPA 21
Db      1 AHGVTSPDTRPAGSTAPPA 21

RESULT 2
AAY96171
ID      AAY96171 standard; peptide; 21 AA.
XX
XX
AC      AAY96171;
XX
XX      19-DEC-2000 (first entry)
XX
XX      MUC1 repeat sequence fragment A1.
DE
XX
XX      MUC1; immunomodulator; glycopeptide; T-lymphocyte; T-cell; proliferation;
KW      cancer; sarcoma; carcinoma; leukaemia; diagnosis; therapy; vaccine;
KW      adjuvant; glycosylation.
XX
XX      Homo sapiens.
OS
XX
XX      Key      Location/Qualifiers
FH      Modified-site 5
FT      /note= "O-glycosylated by GalNAc-beta-1-3Gal"
XX
XX      WO200052046-A1.
PN
XX
XX      08-SEP-2000.
PD
XX
XX      01-MAR-2000; 2000WO-GB000724.
PF
XX
XX      01-MAR-1999; 99GB-00004695.
PR
XX
XX      (IMCR ) IMPERIAL CANCER RES TECHNOLOGY LTD.
PA
XX
XX      Burchell J, Taylor-Papadimitriou J;
PI
XX
XX      WPI; 2000-601868/57.
DR
XX
XX      New immunomodulating glycopeptide that causes super-proliferation of T
PT      cells, useful for treating cells in vitro, for diagnosing or treating
PT      cancer (e.g. carcinoma or sarcoma) or as an adjuvant.
XX
XX      Disclosure; Page 23; 35pp; English.
PS
XX
XX      The present sequence comprises glycopeptide A1, a fragment of the repeat
CC      sequence of MUC1 glycosylated at Thr-5. Glycopeptides comprising this
CC      fragment of MUC1 repeat sequence, especially having a Gal-GalNAc or
CC      GalNAc moiety on Thr-10 or Thr-17 (see AAY96172-74), are useful as
CC      immunomodulators, causing super-proliferation of T cells. Such
CC      glycopeptides can be used in the treatment or diagnosis of a disease, in
CC      particular cancer, or as vaccine adjuvants. The glycopeptides are
CC      particularly useful in manufacturing a medicament for preventing or
CC      treating cancer by stimulating T cells whose receptors recognize the
CC      glycopeptide. They are also useful for diagnosing or treating cancer,
CC      e.g. carcinoma (e.g. mammary, lung, bladder or colon carcinomas, or ovary
CC      and endometrial tumours), or sarcoma (e.g. soft tissue and bone sarcomas,
CC      or leukaemia). Human peripheral blood lymphocytes (PBL) were used in a
CC      proliferation assay. The proliferation index of A1 (taking the index as 1
CC      when no glycopeptide was present) was 0.7-1.9
XX
XX      Sequence 21 AA;
SQ
XX
XX      Query Match      100.0%; Score 113; DB 3; Length 21;
XX      Best Local Similarity 100.0%; Pred. No. 5e-07;
XX      Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AHGVTSPDTRPAGSTAPPA 21
Db      1 AHGVTSPDTRPAGSTAPPA 21

RESULT 4
AAY96169
ID      AAY96169 standard; peptide; 21 AA.
XX
XX      AAY96169;
AC
XX
XX      19-DEC-2000 (first entry)
XX
XX
XX      1 AHGVTSPDTRPAGSTAPPA 21
Db      1 AHGVTSPDTRPAGSTAPPA 21

Query Match      100.0%; Score 113; DB 3; Length 21;
Best Local Similarity 100.0%; Pred. No. 5e-07;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

XX DE MUC1 repeat sequence fragment.
 XX KW MUC1; immunomodulator; glycopeptide; T-lymphocyte; T-cell; proliferation;
 XX KW cancer; sarcoma; carcinoma; leukaemia; diagnosis; therapy; vaccine;
 XX KW adjuvant; glycosylation.
 XX OS Homo sapiens.
 XX PN WO200052046-A1.
 XX PD 08-SEP-2000.
 XX PF 01-MAR-2000; 2000WO-GB000724.
 XX PR 01-MAR-1999; 99GB-00004695.
 XX PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY LTD.
 XX PI Burchell J, Taylor-Papadimitriou J;
 XX DR WPI; 2000-601868/57.
 XX New immunomodulating glycopeptide that causes super-proliferation of T
 PT cells, useful for treating cells in vitro, for diagnosing or treating
 PT cancer (e.g. carcinoma or sarcoma) or as an adjuvant.
 XX Claim 2; Page 3; 35pp; English.
 XX The present sequence comprises a fragment of the repeat sequence of MUC1.
 CC A glycopeptide comprising this sequence, especially having a Gal-GalNac
 CC or GalNac moiety on Thr-10 or Thr-17 (see AA96172-74), is useful as an
 CC immunomodulator causing super-proliferation of T cells. Such
 CC glycopeptides can be used in the treatment or diagnosis of a disease, in
 CC particular cancer, or as a vaccine adjuvant. The glycopeptides are
 CC particularly useful in manufacturing a medicament for preventing or
 CC treating cancer by stimulating T cells whose receptors recognize the
 CC glycopeptide. They are also useful for diagnosing or treating cancer,
 CC e.g. carcinoma (e.g. mammary, lung, bladder or colon carcinomas, or ovary
 CC and endometrial tumours), or sarcoma (e.g. soft tissue and bone sarcomas,
 CC or leukaemia)
 XX SQ Sequence 21 AA;
 Query Match 100.0%; Score 113; DB 3; Length 21;
 Best Local Similarity 100.0%; Pred. No. 5e-07;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AHGVTSA PDTRPAGSTAPPA 21
 DB 1 AHGVTSA PDTRPAGSTAPPA 21
 RESULT 5
 AA96180
 ID AA96180 standard; peptide; 21 AA.
 XX AC AA96180;
 XX DT 19-DEC-2000 (first entry)
 XX DE MUC1 repeat sequence fragment A9.
 XX MUC1; immunomodulator; glycopeptide; T-lymphocyte; T-cell; proliferation;
 KW cancer; sarcoma; carcinoma; leukaemia; diagnosis; therapy; vaccine;
 KW KW adjuvant; glycosylation.
 XX OS Homo sapiens.
 XX Key Location/Qualifiers
 FH Modified-site 5
 FT Modified-site 6 /note= "O-glycosylated by GalNac-beta-1-3Gal"

FT Modified-site 16 /note= "O-glycosylated by GalNac-beta-1-3Gal"
 FT FT Modified-site 17 /note= "O-glycosylated by GalNac-beta-1-3Gal"
 FT FT Modified-site 17 /note= "O-glycosylated by GalNac-beta-1-3Gal"
 XX WO200052046-A1.
 XX PD 08-SEP-2000.
 XX PF 01-MAR-2000; 2000WO-GB000724.
 XX PR 01-MAR-1999; 99GB-00004695.
 XX PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY LTD.
 XX PI Burchell J, Taylor-Papadimitriou J;
 XX DR WPI; 2000-601868/57.
 XX New immunomodulating glycopeptide that causes super-proliferation of T
 PT cells, useful for treating cells in vitro, for diagnosing or treating
 PT cancer (e.g. carcinoma or sarcoma) or as an adjuvant.
 XX Disclosure; Page 23; 35pp; English.
 XX The present sequence comprises glycopeptide A9, a fragment of the repeat
 CC sequence of MUC1 glycosylated at Thr-5, Ser-6, Ser-16 and Thr-17.
 CC Glycopeptides comprising this fragment of MUC1 repeat sequence,
 CC especially having a Gal-GalNac or GalNac moiety on Thr-10 or Thr-17 (see
 CC AA96172-74), are useful as immunomodulators, causing super-proliferation
 CC of T cells. Such glycopeptides can be used in the treatment or diagnosis
 CC of a disease, in particular cancer, or as vaccine adjuvants. The
 CC glycopeptides are particularly useful in manufacturing a medicament for
 CC preventing or treating cancer by stimulating T cells whose receptors
 CC recognize the glycopeptide. They are also useful for diagnosing or
 CC treating cancer, e.g. carcinoma (e.g. mammary, lung, bladder or colon
 CC carcinomas, or ovary and endometrial tumours), or sarcoma (e.g. soft
 CC tissue and bone sarcomas, or leukaemia). Human peripheral blood
 CC lymphocytes were used in a proliferation assay. The proliferation index
 CC of A9 (taking the index as 1 when no glycopeptide was present) was up to
 CC 34
 XX SQ Sequence 21 AA;
 Query Match 100.0%; Score 113; DB 3; Length 21;
 Best Local Similarity 100.0%; Pred. No. 5e-07;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AHGVTSA PDTRPAGSTAPPA 21
 DB 1 AHGVTSA PDTRPAGSTAPPA 21
 RESULT 6
 AA96179
 ID AA96179 standard; peptide; 21 AA.
 XX AC AA96179;
 XX DT 19-DEC-2000 (first entry)
 XX DE MUC1 repeat sequence fragment A8.
 XX MUC1; immunomodulator; glycopeptide; T-lymphocyte; T-cell; proliferation;
 KW cancer; sarcoma; carcinoma; leukaemia; diagnosis; therapy; vaccine;
 KW KW adjuvant; glycosylation.
 XX OS Homo sapiens.
 XX Key Location/Qualifiers
 FH Modified-site 5 /note= "O-glycosylated by GalNac-beta-1-3Gal"
 FT Modified-site 6

```

FT Modified-site 6
FT Modified-site 10 /note= "O-glycosylated by GalNac-beta-1-3Gal"
FT Modified-site 16 /note= "O-glycosylated by GalNac-beta-1-3Gal"
FT Modified-site 17 /note= "O-glycosylated by GalNac-beta-1-3Gal"
FT Modified-site 17 /note= "O-glycosylated by GalNac-beta-1-3Gal"
FT WO200052046-A1.
FN
XX
XX
XX 08-SEP-2000.
XX
XX 01-MAR-2000; 2000WO-GB000724.
XX
XX 01-MAR-1999; 99GB-00004695.
XX
XX (IMCR ) IMPERIAL CANCER RES TECHNOLOGY LTD.
XX
XX Burchell J, Taylor-Papadimitriou J;
XX
XX WPI; 2000-601868/57.
XX
XX New immunomodulating glycopeptide that causes super-proliferation of T
XX cells, useful for treating cells in vitro, for diagnosing or treating
XX cancer (e.g. carcinoma or sarcoma) or as an adjuvant.
XX
XX Disclosure; Page 23; 35pp; English.
XX
XX The present sequence comprises glycopeptide A8, a fragment of the repeat
XX sequence of MUC1 glycosylated at Thr-5, Ser-6, Thr-10, Ser-16 and Thr-17.
XX Glycopeptides comprising this fragment of MUC1 repeat sequence,
XX especially having a Gal-GalNac or GalNac moiety on Thr-10 or Thr-17 (see
XX AAY96172-74), are useful as immunomodulators, causing super-proliferation
XX of T cells. Such glycopeptides can be used in the treatment or diagnosis
XX of a disease, in particular cancer, or as vaccine adjuvants. The
XX glycopeptides are particularly useful in manufacturing a medicament for
XX preventing or treating cancer by stimulating T cells whose receptors
XX recognize the glycopeptide. They are also useful for diagnosing or
XX treating cancer, e.g. carcinoma (e.g. mammary, lung, bladder or colon
XX carcinomas, or ovary and endometrial tumours), or sarcoma (e.g. soft
XX tissue and bone sarcomas, or leukaemia). Human peripheral blood
XX lymphocytes were used in a proliferation assay. The proliferation index
XX of A8 (taking the index as 1 when no glycopeptide was present) was up to
XX 7
XX
XX Sequence 21 AA;
XX
XX Query Match 100.0%; Score 113; DB 3; Length 21;
XX Best Local Similarity 100.0%; Pred. No. 5e-07;
XX Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AHGVTSA PDTRPAGSTAPPA 21
Db |||||
1 AHGVTSA PDTRPAGSTAPPA 21

RESULT 7
AAY96172
ID AAY96172 standard; peptide; 21 AA.
XX
XX AAY96172;
XX
XX 19-DEC-2000 (first entry)
XX
XX MUC1 repeat sequence fragment A2 GalNac.
XX
XX MUC1; immunomodulator; glycopeptide; T-lymphocyte; T-cell; proliferation;
XX cancer; sarcoma; carcinoma; leukaemia; diagnosis; therapy; vaccine;
XX adjuvant; glycosylation.
XX
XX Homo sapiens.
XX

QY 1 AHGVTSA PDTRPAGSTAPPA 21
Db |||||
1 AHGVTSA PDTRPAGSTAPPA 21

RESULT 8
AAY96174
ID AAY96174 standard; peptide; 21 AA.
XX
XX AAY96174;
XX
XX 19-DEC-2000 (first entry)
XX
XX MUC1 repeat sequence fragment A3.
XX
XX MUC1; immunomodulator; glycopeptide; T-lymphocyte; T-cell; proliferation;
XX cancer; sarcoma; carcinoma; leukaemia; diagnosis; therapy; vaccine;
XX adjuvant; glycosylation.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Modified-site 17 /note= "O-glycosylated by GalNac-beta-1-3Gal"
XX
XX WO200052046-A1.
XX

```

```

FH Key Location/Qualifiers
FT Modified-site 10 /note= "O-glycosylated by GalNac"
XX
XX WO200052046-A1.
XX
XX 08-SEP-2000.
XX
XX 01-MAR-2000; 2000WO-GB000724.
XX
XX 01-MAR-1999; 99GB-00004695.
XX
XX (IMCR ) IMPERIAL CANCER RES TECHNOLOGY LTD.
XX
XX Burchell J, Taylor-Papadimitriou J;
XX
XX WPI; 2000-601868/57.
XX
XX New immunomodulating glycopeptide that causes super-proliferation of T
XX cells, useful for treating cells in vitro, for diagnosing or treating
XX cancer (e.g. carcinoma or sarcoma) or as an adjuvant.
XX
XX Disclosure; Page 23; 35pp; English.
XX
XX The present sequence comprises glycopeptide A2 GalNac, a fragment of the
XX repeat sequence of MUC1 glycosylated at Thr-10. Glycopeptides comprising
XX this fragment of MUC1 repeat sequence, especially having a Gal-GalNac or
XX GalNac moiety on Thr-10 or Thr-17 (see AAY96172-74), are useful as
XX immunomodulators, causing super-proliferation of T cells. Such
XX glycopeptides can be used in the treatment or diagnosis of a disease, in
XX particular cancer, or as vaccine adjuvants. The glycopeptides are
XX particularly useful in manufacturing a medicament for preventing or
XX treating cancer by stimulating T cells whose receptors recognize the
XX glycopeptide. They are also useful for diagnosing or treating cancer,
XX e.g. carcinoma (e.g. mammary, lung, bladder or colon carcinomas, or ovary
XX and endometrial tumours), or sarcoma (e.g. soft tissue and bone sarcomas,
XX or leukaemia). Human peripheral blood lymphocytes (PBL) were used in a
XX proliferation assay. The proliferation index of A2 GalNac (taking the
XX index as 1 when no glycopeptide was present) was 48
XX
XX Sequence 21 AA;
XX
XX Query Match 100.0%; Score 113; DB 3; Length 21;
XX Best Local Similarity 100.0%; Pred. No. 5e-07;
XX Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AHGVTSA PDTRPAGSTAPPA 21
Db |||||
1 AHGVTSA PDTRPAGSTAPPA 21

RESULT 8
AAY96174
ID AAY96174 standard; peptide; 21 AA.
XX
XX AAY96174;
XX
XX 19-DEC-2000 (first entry)
XX
XX MUC1 repeat sequence fragment A3.
XX
XX MUC1; immunomodulator; glycopeptide; T-lymphocyte; T-cell; proliferation;
XX cancer; sarcoma; carcinoma; leukaemia; diagnosis; therapy; vaccine;
XX adjuvant; glycosylation.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Modified-site 17 /note= "O-glycosylated by GalNac-beta-1-3Gal"
XX
XX WO200052046-A1.
XX

```


PI Burchell J, Taylor-Papadimitriou J;
 XX WPI; 2000-601868/57.
 XX
 XX
 PT New immunomodulating glycopeptide that causes super-proliferation of T
 PT cells, useful for treating cells in vitro, for diagnosing or treating
 PT cancer (e.g. carcinoma or sarcoma) or as an adjuvant.
 XX
 XX
 PS Disclosure; Page 23; 35pp; English.
 XX
 XX The present sequence comprises glycopeptide A7, a fragment of the repeat
 CC sequence of MUC1 glycosylated at Thr-5, Ser-16 and Thr-17. Glycopeptides
 CC comprising this fragment of MUC1 repeat sequence, especially having a Gal
 CC -GalNAc or GalNAc moiety on Thr-10 or Thr-17 (see AA96172-74), are
 CC useful as immunomodulators, causing super-proliferation of T cells. Such
 CC glycopeptides can be used in the treatment or diagnosis of a disease, in
 CC particular cancer, or as vaccine adjuvants. The glycopeptides are
 CC particularly useful in manufacturing T cells whose receptors recognize the
 CC treating cancer by stimulating T cells whose receptors recognizing the
 CC glycopeptide. They are also useful for diagnosing or treating cancer,
 CC e.g. carcinoma (e.g. mammary, lung, bladder or colon carcinomas, or ovary
 CC and endometrial tumours), or sarcoma (e.g. soft tissue and bone sarcomas,
 CC or leukaemia). Human peripheral blood lymphocytes were used in a
 CC proliferation assay. The proliferation index of A7 (taking the index as 1
 CC when no glycopeptide was present) was up to 13
 XX
 XX Sequence 21 AA;
 SQ
 Query Match 100.0%; Score 113; DB 3; Length 21;
 Best Local Similarity 100.0%; Pred. No. 5e-07;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AHGVTSPDTRPAPGSTAPPA 21
 Db 1 AHGVTSPDTRPAPGSTAPPA 21
 ||||||||||||||||||
 RESULT 11
 AA96177
 ID AAY96177 standard; peptide; 21 AA.
 XX
 AC AAY96177;
 XX
 DT 19-DEC-2000 (first entry)
 XX
 DE MUC1 repeat sequence fragment A6.
 XX
 KW MUC1; immunomodulator; glycopeptide; T-lymphocyte; T-cell; proliferation;
 KW cancer; sarcoma; carcinoma; leukaemia; diagnosis; therapy; vaccine;
 KW adjuvant; glycosylation.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Modified-site 5 /note= "O-glycosylated by GalNAc-beta-1-3Gal"
 FT Modified-site 17
 FT Modified-site 17 /note= "O-glycosylated by GalNAc-beta-1-3Gal"
 XX
 PN WO200052046-A1.
 XX
 PD 08-SEP-2000.
 XX
 PF 01-MAR-2000; 2000WO-GB000724.
 XX
 PR 01-MAR-1999; 99GB-00004695.
 XX
 XX (IMCR) IMPERIAL CANCER RES TECHNOLOGY LTD.
 XX Burchell J, Taylor-Papadimitriou J;
 XX WPI; 2000-601868/57.
 XX
 XX
 PT New immunomodulating glycopeptide that causes super-proliferation of T
 PT cells, useful for treating cells in vitro, for diagnosing or treating
 PT cancer (e.g. carcinoma or sarcoma) or as an adjuvant.
 XX
 XX
 PS Disclosure; Page 23; 35pp; English.
 XX
 XX The present sequence comprises glycopeptide A6, a fragment of the repeat
 CC sequence of MUC1 glycosylated at Thr-5 and Thr-17. Glycopeptides
 CC comprising this fragment of MUC1 repeat sequence, especially having a Gal
 CC -GalNAc or GalNAc moiety on Thr-10 or Thr-17 (see AA96172-74), are
 CC useful as immunomodulators, causing super-proliferation of T cells. Such
 CC glycopeptides can be used in the treatment or diagnosis of a disease, in
 CC particular cancer, or as vaccine adjuvants. The glycopeptides are
 CC particularly useful in manufacturing T cells whose receptors recognize the
 CC treating cancer by stimulating T cells whose receptors recognizing the
 CC glycopeptide. They are also useful for diagnosing or treating cancer,
 CC e.g. carcinoma (e.g. mammary, lung, bladder or colon carcinomas, or ovary
 CC and endometrial tumours), or sarcoma (e.g. soft tissue and bone sarcomas,
 CC or leukaemia). Human peripheral blood lymphocytes were used in a
 CC proliferation assay. The proliferation index of A6 (taking the index as 1
 CC when no glycopeptide was present) was up to 34
 XX
 XX Sequence 21 AA;
 SQ
 Query Match 100.0%; Score 113; DB 3; Length 21;
 Best Local Similarity 100.0%; Pred. No. 5e-07;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AHGVTSPDTRPAPGSTAPPA 21
 Db 1 AHGVTSPDTRPAPGSTAPPA 21
 ||||||||||||||||||
 RESULT 12
 AA96175
 ID AAY96175 standard; peptide; 21 AA.
 XX
 AC AAY96175;
 XX
 DT 19-DEC-2000 (first entry)
 XX
 DE MUC1 repeat sequence fragment A4.
 XX
 KW MUC1; immunomodulator; glycopeptide; T-lymphocyte; T-cell; proliferation;
 KW cancer; sarcoma; carcinoma; leukaemia; diagnosis; therapy; vaccine;
 KW adjuvant; glycosylation.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Modified-site 6 /note= "O-glycosylated by GalNAc-beta-1-3Gal"
 FT Modified-site 6
 FT Modified-site 6
 XX
 PN WO200052046-A1.
 XX
 PD 08-SEP-2000.
 XX
 PF 01-MAR-2000; 2000WO-GB000724.
 XX
 PR 01-MAR-1999; 99GB-00004695.
 XX
 XX (IMCR) IMPERIAL CANCER RES TECHNOLOGY LTD.
 XX Burchell J, Taylor-Papadimitriou J;
 XX WPI; 2000-601868/57.
 XX
 XX
 PT New immunomodulating glycopeptide that causes super-proliferation of T
 PT cells, useful for treating cells in vitro, for diagnosing or treating
 PT cancer (e.g. carcinoma or sarcoma) or as an adjuvant.
 XX
 XX
 PS Disclosure; Page 23; 35pp; English.
 XX
 XX

CC The present sequence comprises glycopeptide A4, a fragment of the repeat
 CC sequence of MUC1 glycosylated at Ser-6. Glycopeptides comprising this
 CC fragment of MUC1 repeat sequence, especially having a Gal-GalNAc or
 CC GalNAc moiety on Thr-10 or Thr-17 (see AAY96172-74), are useful as
 CC immunomodulators, causing super-proliferation of T cells. Such
 CC glycopeptides can be used in the treatment or diagnosis of a disease, in
 CC particular cancer, or as vaccine adjuvants. The glycopeptides are
 CC particularly useful in manufacturing a medicament for preventing or
 CC treating cancer by stimulating T cells whose receptors recognize the
 CC glycopeptide. They are also useful for diagnosing or treating cancer,
 CC e.g. carcinoma (e.g. mammary, lung, bladder or colon carcinomas, or ovary
 CC and endometrial tumours), or sarcoma (e.g. soft tissue and bone sarcomas,
 CC or leukaemia). Human peripheral blood lymphocytes (PBL) were used in a
 CC proliferation assay. The proliferation index of A4 (taking the index as 1
 CC when no glycopeptide was present) was up to 35

XX Sequence 21 AA;

Query Match 100.0%; Score 113; DB 3; Length 21;
 Best Local Similarity 100.0%; Pred. No. 5e-07;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AHGVTSA PDTRPAGSTAPPA 21

DB 1 AHGVTSA PDTRPAGSTAPPA 21

RESULT 13

ADK65929

ID ADK65929 standard; peptide; 21 AA.

XX

AC ADK65929;

DT 06-MAY-2004 (first entry)

XX MUC1 glycopeptide.

DE

XX glycopeptide; immunostimulating mucin; MUC1; immune response; cancer;
 KW antibody.

XX Unidentified.

XX Key Location/Qualifiers
 FT Modified-site 10.11
 FT /note= "modified by GalNAc"

XX

XX WO2004009632-A2.

XX

XX 29-JAN-2004.

XX

XX 22-JUL-2003; 2003WO-EP008014.

XX

XX 22-JUL-2002; 2002EP-00016440.

XX

XX (NEMO-) NEMO IMMUNOTHERAPIE AG.

XX

XX Goletz S, Karsten U;

XX WPI; 2004-123378/12.

XX

XX Production of mucin, MUC1, molecules that induce an immune response,
 PT useful for diagnosis, prevention and treatment of tumors, by selection
 PT for specific binding to antibodies.

XX Example 3; Page 47; 83pp; German.

XX

XX The present invention relates to a method for producing an
 CC immunostimulating mucin MUC1 molecule that can induce an immune response
 CC in humans. MUC1, also cells and cell lysates containing it, or antibodies
 CC that recognize it, are useful for treatment, prevention, diagnosis
 CC (including in vivo), monitoring and prognosis of tumours, particularly
 CC for treatment of minimal residual disease or metastases. The present
 CC sequence is a MUC1 glycopeptide.

XX Sequence 21 AA;

Query Match 100.0%; Score 113; DB 8; Length 21;
 Best Local Similarity 100.0%; Pred. No. 5e-07;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AHGVTSA PDTRPAGSTAPPA 21

DB 1 AHGVTSA PDTRPAGSTAPPA 21

RESULT 14

ADO58854

ID ADO58854 standard; peptide; 21 AA.

XX

AC ADO58854;

DT 15-JUL-2004 (first entry)

XX Human MUC1 immunogene glycopeptide tandem-repeat domain fragment #1.

XX tandem-repeat domain; TRD; MUC1; vaccine; MUC1-expressing cancer; cancer;
 KW breast carcinoma; colorectal carcinoma; pancreatic carcinoma;
 KW gastric carcinoma; immunogene; glycopeptide; human.

XX Homo sapiens.

XX DE10305607-A1.

XX 18-MAR-2004.

XX

XX 11-FEB-2003; 2003DE-01005607.

XX 05-SEP-2002; 2002DE-01041207.

XX (CELL-) CELL CENT COLOGNE GMBH.

XX Hanisch F;

XX WPI; 2004-249134/24.

XX

XX New peptide fragments from the tandem repeat domain of the MUC1 protein,
 PT useful for preparing therapeutic compositions, such as, vaccines for
 PT treatment of cancer.

XX Claim 8; SEQ ID NO 5; 26pp; German.

XX

XX The invention relates to peptide fragments from the tandem-repeat domain
 CC (TRD) of the MUC1 protein. The peptide fragments and also antigen
 CC presenting cells that present them, are useful in therapeutic
 CC compositions, including vaccines, for treatment of MUC1-expressing
 CC cancers, particularly breast, colorectal, pancreatic and gastric
 CC carcinoma. The present sequence represents a human MUC1 immunogene
 CC glycopeptide tandem-repeat domain fragment.

XX Sequence 21 AA;

Query Match 100.0%; Score 113; DB 8; Length 21;
 Best Local Similarity 100.0%; Pred. No. 5e-07;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AHGVTSA PDTRPAGSTAPPA 21

DB 1 AHGVTSA PDTRPAGSTAPPA 21

RESULT 15

ADC78317

ID ADC78317 standard; peptide; 25 AA.

XX

AC ADC78317;

XX

Tue Apr 18 08:15:23 2006

us-09-606-910e-1.rag

```
DT 01-JAN-2004 (first entry)
XX
DE MUC1 peptide TAP2 used to immunoaffinity-purify MUC1 antibodies.
XX
KW cancer associated marker protein; carcinogenic; MUC1; TAP2.
XX
OS Unidentified.
XX
FH Key Location/Qualifiers
FT Modified-site 9 /label= OTHER
FT /note= "OTHER = Residue is O-glycosylated with N-acetyl-
FT galactosamine"
FT Modified-site 21
FT /label= OTHER
FT /note= "OTHER = Residue is O-glycosylated with N-acetyl-
FT galactosamine"
XX
PN WO200034787-A1.
XX
PD 15-JUN-2000.
XX
PF 10-DEC-1999; 99WO-GB004182.
XX
PR 10-DEC-1998; 98GB-00027228.
XX
PA (UYNO-) UNIV NOTTINGHAM.
XX
PI Robertson JR, Graves CRL, Price MR;
XX
DR WPI; 2000-431380/37.
XX
PT Detecting presence of cancer associated marker proteins for detecting and
PT monitoring cancer involves contacting the sample with auto antibodies
PT against the marker protein from the same species as the sample.
XX
PS Example 4; Fig 4; 41pp; English.
XX
CC The invention relates to a novel method for detecting cancer associated
CC marker proteins in mammals which involves contacting the body fluid with
CC mammalian autoantibodies against at least one epitope of a cancer
CC associated marker protein from the same species as the sample and
CC detecting the antibody-cancer associated marker protein complex formed.
CC The method of the invention may be useful for detecting cancers and
CC screening for recurrence of cancer after treatment, monitoring systemic
CC therapies and selecting appropriate therapies. The method may also be
CC useful for assessment of predisposition of an individual to cancer,
CC detection of pre-neoplastic or carcinogenic modifications in asymptomatic
CC patients, for diagnosing primary and secondary cancers and monitoring the
CC progression of cancer. The current sequence is that of the MUC1 peptide
CC TAP2 of the invention which was used to immunoaffinity-purify MUC1
CC antibodies.
XX
SQ Sequence 25 AA;
Query Match 100.0%; Score 113; DB 3; Length 25;
Best Local Similarity 100.0%; Pred. No. 5.9e-07;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AHGVTSA PDTRPAPGSTAPPA 21
Db 5 AHGVTSA PDTRPAPGSTAPPA 25
|||||
Search completed: April 14, 2006, 10:24:14
Job time : 318.427 secs
```

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 14, 2006, 18:45:20 ; Search time 1733.78 Seconds

(without alignments)
688.503 Million cell updates/sec

Title: US-09-606-910E-1

Perfect score: 113
Sequence: 1 AHGVTAPDTPRPGSTAPPA 21

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame_p2n.model -DEV=xlp
-Q=/abs/ABSWEB/spool/US09606910/runat_14042006_091619_1514/app_query.fasta_1
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -NORM=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abes05p
-USER=US09606910 @CGN 1 1 7724 @runat_14042006_091619_1514 -NCPU=6 -ICPU=3
-NO MMAP -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*

- 1: gb.ba.*
- 2: gb.in.*
- 3: gb.env.*
- 4: gb.cm.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pr.*
- 9: gb.ro.*
- 10: gb.sts.*
- 11: gb.sy.*
- 12: gb.un.*
- 13: gb.vi.*
- 14: gb.htg.*
- 15: gb.pl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	113	100.0	120	6	AX192396 Sequence
C 2	113	100.0	120	6	BD000571 Human pol
C 3	113	100.0	525	6	BD225141 Medicinal

4	113	100.0	891	6	BD225147
5	113	100.0	1125	6	CS085358
6	113	100.0	1371	6	BD225144
7	113	100.0	1414	11	AF423031
8	113	100.0	1455	6	CQ715242
9	113	100.0	1457	6	AX959914
10	113	100.0	1572	6	AX093798
11	113	100.0	1721	6	CQ771290
12	113	100.0	1721	6	AR492306
13	113	100.0	1721	6	AX335860
14	113	100.0	1721	6	AX440427
15	113	100.0	1721	6	AX587588
16	113	100.0	1721	8	HSTEVMA
17	113	100.0	1737	6	BD225150
18	113	100.0	1774	6	AX959684
19	113	100.0	1774	6	AX959912
20	113	100.0	1800	6	CQ875507
21	113	100.0	1804	6	AR492318
22	113	100.0	1804	6	AX335367
23	113	100.0	1804	8	HUMMUCAB
24	113	100.0	1818	6	CQ875500
25	113	100.0	1818	6	CQ875505
26	113	100.0	1818	6	CS057842
27	113	100.0	1834	11	AF423030
28	113	100.0	1835	6	AX959918
29	113	100.0	1857	6	CS057851
30	113	100.0	1857	6	CS057852
31	113	100.0	1878	6	CS057850
32	113	100.0	1971	6	AX963157
33	113	100.0	2037	6	AX963159
34	113	100.0	2135	6	AX959916
35	113	100.0	2238	8	HSEETA
36	113	100.0	2297	6	BD272907
37	113	100.0	2487	6	CS047899
38	113	100.0	3670	6	CS056288
39	113	100.0	3694	6	CS056284
40	113	100.0	3712	6	CS056290
41	113	100.0	3727	6	CS056287
42	113	100.0	3769	6	CS056289
43	113	100.0	4003	6	CS047901
44	113	100.0	4139	6	CQ834017
45	113	100.0	4139	6	CS111465

ALIGNMENTS

RESULT 1	AX192396/c	AX192396	Sequence 2 from Patent EP1103623.	120 bp	DNA	linear	PAT 15-AUG-2001
LOCUS	AX192396	AX192396	AX192396.1	GI:15210363			
DEFINITION	Sequence 2 from Patent EP1103623.						
ACCESSION	AX192396						
VERSION	AX192396.1						
KEYWORDS	Homo sapiens (human)						
SOURCE	Homo sapiens						
ORGANISM	Homo sapiens						
REFERENCE	1						
AUTHORS	Taylor-Papadimitriou,J., Burchell,J. and Gendler,S.						
TITLE	Human mucin core protein: nucleic acid probes, peptide fragments and antibodies thereto, and uses thereof in diagnostic and therapeutic methods						
JOURNAL	Patent: EP 1103623-A 2 30-MAY-2001;						
FEATURES	IMPERIAL CANCER RESEARCH TECHNOLOGY LIMITED (GB)						
source	1. .120						
	/organism="Homo sapiens"						
	/mol_type="unassigned DNA"						
	/db_xref="taxon:9606"						

ORIGIN

Alignment Scores:

```
Pred. No.: 0.00167 Length: 120
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservatives: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 6 Gaps: 0

US-09-606-910E-1 (1-21) x AX192396 (1-120)

QY 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db 69 GCCACGGTGTACCTCGGCCCGGACACAGCGCGCGGCTCCACCGCCCCCA 10

QY 21 Ala 21
Db 9 GCC 7

RESULT 2
LOCUS BD000571/c 120 bp DNA linear PAT 31-JAN-2002
DEFINITION Human polymorphic epithelial mucin core protein and nucleic acid
encoding the protein.
ACCESSION BD000571
VERSION BD000571.1 GI:18623684
KEYWORDS JP 2000333675-A/2.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.
REFERENCE 1 (bases 1 to 120)
AUTHORS Papadimitrov,J.T., Jendora,S. and Bachieru,J.
TITLE Human polymorphic epithelial mucin core protein and nucleic acid
encoding the protein
JOURNAL Patent: JP 2000333675-A 2 05-DEC-2000;
IMPERIAL CANCER RESEARCH TECHNOLOGY LTD
COMMENT OS Homo sapiens (human)
PN JP 2000333675-A/2
PD 05-DEC-2000
PF 26-APR-2000 JP 2000125724
PR 07-JAN-1987 GB 8700269,07-JAN-1987 GB 8700279 PR
22-APR-1987 US 041306,09-NOV-1987 GB 8726172 PI JOYCE
TAYLOR PAPADIMITROV,SANDRA JENDORA,JOY BACHIERU PC
C12N15/02,A61K38/00,A61K39/395,A61K39/395,A61K49/00,A61P35/00, PC
C07K14/47,
PC C07K16/44,C12N5/10,C12P21/08//(C12N15/02,C12R1:91),(C12N5/10,
PC C12R1:91),
PC C12N15/00,A61K37/02,C12N5/00,(C12N15/00,C12R1:91),(C12N5/00,
PC C12R1:91)
CC
FH Key Location/Qualifiers
FT source 1..120
FT 1..120 /organism='Homo sapiens (human)'.
FEATURES
source
1..120
/organism='Homo sapiens'
/mol_type='genomic DNA'
/db_xref='taxon:9606'

ORIGIN
Alignment Scores:
Pred. No.: 0.00167 Length: 120
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservatives: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 6 Gaps: 0

US-09-606-910E-1 (1-21) x BD000571 (1-120)

QY 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db 69 GCCACGGTGTACCTCGGCCCGGACACAGCGCGCGGCTCCACCGCCCCCA 10

Pred. No.: 0.00167 Length: 120
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservatives: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 6 Gaps: 0

US-09-606-910E-1 (1-21) x AX192396 (1-120)

QY 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db 69 GCCACGGTGTACCTCGGCCCGGACACAGCGCGCGGCTCCACCGCCCCCA 10

QY 21 Ala 21
Db 9 GCC 7

RESULT 3
LOCUS BD225141 525 bp DNA linear PAT 17-JUL-2003
DEFINITION Medicinal composition having antitumor effect and containing DNA
encoding antigenic protein.
ACCESSION BD225141
VERSION BD225141.1 GI:33034911
KEYWORDS JP 2002528519-A/2.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.
REFERENCE 1 (bases 1 to 525)
AUTHORS Pallente,D., Massimo,A.M.D. and Desantis,R.
TITLE Medicinal composition having antitumor effect and containing DNA
encoding antigenic protein
JOURNAL Patent: JP 2002528519-A 2 03-SEP-2002;
MENARINI RICERCHE SPA
COMMENT OS Homo sapiens (human)
PN JP 2002528519-A/2
PD 03-SEP-2002
PF 18-OCT-1999 JP 2000579265
PR 30-OCT-1998 IT M198A002330
PI DINO PALLENTE,ANNA MARIA D MASSIMO,RITA DESANTIS PC
A61K38/00,A61K35/76,A61K39/00,A61K48/00,A61P35/00,C12N15/09, PC
A61K37/02.
PC C12N15/00
CC Medicinal composition having antitumor effect and containing
DNA encoding
antigenic protein
FH Key Location/Qualifiers
FT source 1..525
FT 1..525 /organism='Homo sapiens (human)'.
FEATURES
source
1..525
/organism='Homo sapiens'
/mol_type='genomic DNA'
/db_xref='taxon:9606'

ORIGIN
Alignment Scores:
Pred. No.: 0.00558 Length: 525
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservatives: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 6 Gaps: 0

US-09-606-910E-1 (1-21) x BD225141 (1-525)

QY 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db 274 GCACACGGGTGTACTCGGCTCCGGATACCGCGCGCCAGGTAGTACCGCCCTCT 333

QY 21 Ala 21
Db 334 GCC 336

RESULT 4
LOCUS BD225147 891 bp DNA linear PAT 17-JUL-2003
DEFINITION Medicinal composition having antitumor effect and containing DNA
encoding antigenic protein.
ACCESSION BD225147
VERSION BD225147.1 GI:33034917
KEYWORDS JP 2002528519-A/8.
SOURCE Homo sapiens (human)
```

```
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
1 (bases 1 to 891)
REFERENCE
AUTHORS Pallente,D., Massimo,A.M.D. and Desantis,R.
TITLE Medicinal composition having antitumor effect and containing DNA
encoding antigenic protein
JOURNAL Patent: JP 2002528519-A 8 03-SEP-2002;
MENARINI RICERCHE SPA
COMMENT OS Homo sapiens (human)
PN JP 2002528519-A/8
PD 03-SEP-2002
PF 18-OCT-1999 JP 2000579265
PI 30-OCT-1998 IT M198A002330
PI DINO PALLEENTE, ANNA MARIA D MASSIMO, RITA DESANTIS PC
A61K38/00, A61K35/76, A61K39/00, A61K48/00, A61P35/00, C12N15/09, PC
A61K37/02,
PC C12N15/00
CC Medicinal composition having antitumor effect and containing
CC antigenic protein
FH Key Location/Qualifiers
FT source 1..891
FT /organism='Homo sapiens (human)'.
FEATURES
source
Location/Qualifiers
1..891
/organism='Homo sapiens'
/mol_type='genomic DNA'
/db_xref='taxon:9606'
ORIGIN
Alignment Scores:
Pred. No.: 0.00859 Length: 891
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 6 Gaps: 0
US-09-606-910E-1 (1-21) x BD225147 (1-891)
QY 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
|||||
Db 640 GCACACGGTGTACCTCGGATACAGCGCGCCCGAGGTAGTACCGCCCTCT 699
QY 21 Ala 21
Db 700 GCC 702
RESULT 5
CS085358 1125 bp DNA linear PAT 25-MAY-2005
LOCUS Sequence 14 from Patent WO2005042573.
DEFINITION CS085358
ACCESSION CS085358
VERSION CS085358.1 GI:66711285
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
1
REFERENCE
AUTHORS Kharbada, S. and Kufe, D.W.
TITLE Modulation of the interaction of mucl1 with mucl1 ligands
JOURNAL Patent: WO 2005042573-A 14 12-MAY-2005;
Dana-Farber Cancer Institute, Inc. (US); Ilex Products, Inc. (US)
FEATURES
source
Location/Qualifiers
1..1125
/organism='Homo sapiens'
/mol_type='unassigned DNA'
/db_xref='taxon:9606'
ORIGIN
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
1 (bases 1 to 891)
REFERENCE
AUTHORS Pallente,D., Massimo,A.M.D. and Desantis,R.
TITLE Medicinal composition having antitumor effect and containing DNA
encoding antigenic protein
JOURNAL Patent: JP 2002528519-A 8 03-SEP-2002;
MENARINI RICERCHE SPA
COMMENT OS Homo sapiens (human)
PN JP 2002528519-A/8
PD 03-SEP-2002
PF 18-OCT-1999 JP 2000579265
PI 30-OCT-1998 IT M198A002330
PI DINO PALLEENTE, ANNA MARIA D MASSIMO, RITA DESANTIS PC
A61K38/00, A61K35/76, A61K39/00, A61K48/00, A61P35/00, C12N15/09, PC
A61K37/02,
PC C12N15/00
CC Medicinal composition having antitumor effect and containing
CC antigenic protein
FH Key Location/Qualifiers
FT source 1..891
FT /organism='Homo sapiens (human)'.
FEATURES
source
Location/Qualifiers
1..891
/organism='Homo sapiens'
/mol_type='genomic DNA'
/db_xref='taxon:9606'
ORIGIN
Alignment Scores:
Pred. No.: 0.00859 Length: 891
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 6 Gaps: 0
US-09-606-910E-1 (1-21) x BD225147 (1-891)
QY 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
|||||
Db 640 GCACACGGTGTACCTCGGATACAGCGCGCCCGAGGTAGTACCGCCCTCT 699
QY 21 Ala 21
Db 700 GCC 702
RESULT 5
CS085358 1125 bp DNA linear PAT 25-MAY-2005
LOCUS Sequence 14 from Patent WO2005042573.
DEFINITION CS085358
ACCESSION CS085358
VERSION CS085358.1 GI:66711285
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
1
REFERENCE
AUTHORS Kharbada, S. and Kufe, D.W.
TITLE Modulation of the interaction of mucl1 with mucl1 ligands
JOURNAL Patent: WO 2005042573-A 14 12-MAY-2005;
Dana-Farber Cancer Institute, Inc. (US); Ilex Products, Inc. (US)
FEATURES
source
Location/Qualifiers
1..1125
/organism='Homo sapiens'
/mol_type='unassigned DNA'
/db_xref='taxon:9606'
ORIGIN
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
1 (bases 1 to 1371)
REFERENCE
AUTHORS Pallente,D., Massimo,A.M.D. and Desantis,R.
TITLE Medicinal composition having antitumor effect and containing DNA
encoding antigenic protein
JOURNAL Patent: JP 2002528519-A 5 03-SEP-2002;
MENARINI RICERCHE SPA
COMMENT OS Homo sapiens (human)
PN JP 2002528519-A/5
PD 03-SEP-2002
PF 18-OCT-1999 JP 2000579265
PI 30-OCT-1998 IT M198A002330
PI DINO PALLEENTE, ANNA MARIA D MASSIMO, RITA DESANTIS PC
A61K38/00, A61K35/76, A61K39/00, A61K48/00, A61P35/00, C12N15/09, PC
A61K37/02,
PC C12N15/00
CC Medicinal composition having antitumor effect and containing
CC antigenic protein
FH Key Location/Qualifiers
FT source 1..1371
FT /organism='Homo sapiens (human)'.
FEATURES
source
Location/Qualifiers
1..1371
/organism='Homo sapiens'
/mol_type='genomic DNA'
/db_xref='taxon:9606'
ORIGIN
Alignment Scores:
Pred. No.: 0.0122 Length: 1371
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 6 Gaps: 0
US-09-606-910E-1 (1-21) x BD225144 (1-1371)
QY 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
|||||
Db 403 GCCACGGTGTACCTCGGATACAGCGCGCCCGAGGTAGTACCGCCCTCT 462
QY 21 Ala 21
Db 463 GCC 465
RESULT 6
BD225144 1371 bp DNA linear PAT 17-JUL-2003
LOCUS Medicinal composition having antitumor effect and containing DNA
DEFINITION encoding antigenic protein.
ACCESSION BD225144
VERSION BD225144.1 GI:33034914
KEYWORDS JP 2002528519-A/5.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
1 (bases 1 to 1371)
REFERENCE
AUTHORS Pallente,D., Massimo,A.M.D. and Desantis,R.
TITLE Medicinal composition having antitumor effect and containing DNA
encoding antigenic protein
JOURNAL Patent: JP 2002528519-A 5 03-SEP-2002;
MENARINI RICERCHE SPA
COMMENT OS Homo sapiens (human)
PN JP 2002528519-A/5
PD 03-SEP-2002
PF 18-OCT-1999 JP 2000579265
PI 30-OCT-1998 IT M198A002330
PI DINO PALLEENTE, ANNA MARIA D MASSIMO, RITA DESANTIS PC
A61K38/00, A61K35/76, A61K39/00, A61K48/00, A61P35/00, C12N15/09, PC
A61K37/02,
PC C12N15/00
CC Medicinal composition having antitumor effect and containing
CC antigenic protein
FH Key Location/Qualifiers
FT source 1..1371
FT /organism='Homo sapiens (human)'.
FEATURES
source
Location/Qualifiers
1..1371
/organism='Homo sapiens'
/mol_type='genomic DNA'
/db_xref='taxon:9606'
ORIGIN
Alignment Scores:
Pred. No.: 0.0122 Length: 1371
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 6 Gaps: 0
US-09-606-910E-1 (1-21) x BD225144 (1-1371)
QY 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
|||||
Db 343 GCACACGGTGTACCTCGGATACAGCGCGCCCGAGGTAGTACCGCCCTCT 402
```

```

QY      21 Ala 21
Db      403 GCC 405

RESULT 7
LOCUS   AF423031      1414 bp      mRNA      linear      SYN 10-JUL-2003
DEFINITION Synthetic construct Homo sapiens mucin variant MUC1-CT58 (MUC1)
ACCESSION AF423031
VERSION   AF423031.1 GI:19338621
KEYWORDS .
SOURCE   synthetic construct
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1414)
AUTHORS Hinojosa-Kurtzberg,A.M., Johansson,M.E., Madsen,C.S., Hansson,G.C.
and Gendler,S.J.
TITLE Novel MUC1 splice variants contribute to mucin overexpression in
CFTR-deficient mice
JOURNAL Am. J. Physiol. Gastrointest. Liver Physiol. 284 (5), G853-G862
(2003)
PUBMED 12529261
REFERENCE 2 (bases 1 to 1414)
AUTHORS Hinojosa-Kurtzberg,A.M. and Gendler,S.J.
TITLE Direct Submission
JOURNAL Submitted (21-SEP-2001) Biochemistry, Mayo Clinic, 13400 E. Shea
Boulevard, Scottsdale, AZ 85259, USA
FEATURES
source
1. .1414
/organism="synthetic construct"
/mol_type="mRNA"
/db_xref="taxon:32630"
/note="Homo sapiens gene in transgenic Mus musculus
C57BL/6; isolated from intestinal mucosa"
gene 1. .1414
/organism="MUC1"
CDS 1. .1386
/organism="MUC1"
/note="alternatively spliced; contains exon 6b resulting
in variant carboxy-terminal domain; lacks sites for
beta-catenin and Grb2 interactions; derived from Homo
sapiens"
/codon_start=1
/transl_table=11
/product="mucin variant MUC1-CT58"
/protein_id="AAL86735.1"
/db_xref="GI:19338622"
/translation="MTRPGTQSPFFLLLLTLVTVTGSHASSTPGKEKTSATORSS
VRSSEKKAIVSMTSVLSSHSPGSGSSSTQGQDVLAPATPEPASGAATWGDQVTSVP
VTRPALGSTTPPAHDVTSAPDNKPAAGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS
APDNPALGSTAPPVHVNTSASGSASGLVHNGTSARATTPASKSTPFSPFHH
SDPTTTLASHSTKTDASTHSTHSTVPLTSSNHSSTPQLSTGVSPFPLHINLQFNS
SLEDSTDYQELORDISEMFLQIVKQGGFLGSLNFKRPGSVVQVLTAPREGTINV
HDEVTQFNQYKTEAARVNLTIISDVUSVDVPPFSAQSGAGVPGMGIALLVLCVLA
LAIVLIALAVQCRRKNYQQLDIFPARDTHPMSEYPTHTHGRVYVPPSPSTRSPYE
KEWRVDREKLA"

ORIGIN
Alignment Scores:
Pred. No.: 0.0125 Length: 1414
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 11 Gaps: 0

US-09-606-910E-1 (1-21) x AF423031 (1-1414)

QY      1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db      403 GCCACGGGTGTCACTCGCGCCGGACACACAGCGCGCGCGCGGTCCACCGCCCCCA 462

RESULT 8
LOCUS   CQ715242      1455 bp      DNA      linear      PAT 03-FEB-2004
DEFINITION Sequence 1176 from Patent WO02068579.
ACCESSION CQ715242
VERSION   CQ715242.1 GI:42276099
KEYWORDS .
SOURCE   Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Venter,C.J., Adams,M.C., Li,P.W. and Myers,E.W.
TITLE Kits, such as nucleic acid arrays, comprising a majority of
humanexons or transcripts, for detecting expression and other uses
thereof
JOURNAL Patent: WO 02068579-A 1176 06-SEP-2002;
PE Corporation (NY) (US)
FEATURES
source
1. .1455
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Alignment Scores:
Pred. No.: 0.0128 Length: 1455
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 6 Gaps: 0

US-09-606-910E-1 (1-21) x CQ715242 (1-1455)

QY      1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db      430 GCCACGGGTGTCACTCGCGCCGGACACACAGCGCGCGCGGTCCACCGCCCCCA 489

RESULT 9
LOCUS   AX959914      1457 bp      DNA      linear      PAT 14-JAN-2004
DEFINITION Sequence 19 from Patent WO03100060.
ACCESSION AX959914
VERSION   AX959914.1 GI:40880143
KEYWORDS .
SOURCE   Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Burden,N.G., Ellis,J.H. and Hamblin,P.A.
TITLE Muc-1 antigen with reduced number of vntr repeat units
JOURNAL Patent: WO 03100060-A 19 04-DEC-2003;
GLAXO GROUP LIMITED (GB)
FEATURES
source
1. .1457
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

```



```
US-09-606-910E-1 (1-21) x AR492306 (1-1721)

QY 1 AlaHisGlyValThrSerAlaProGlyThrArgProAlaProGlySerThrAlaProPro 20
Db 580 GCCCAGGGTGTACCTCGGCCCGGACACAGCGCGGCCCGGGCTCCACCGCCCCCA 639

QY 21 Ala 21
Db 640 GCC 642

RESULT 13
AX440427
LOCUS AX335860 1721 bp DNA linear PAT 09-JAN-2002
DEFINITION Sequence 6369 from Patent WO0194629.
ACCESSION AX335860
VERSION AX335860.1 GI:18126579
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1 Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,
Horrihan, S., Soppet, D.R. and Weaver, Z.
TITLE Cancer gene determination and therapeutic screening using signature
gene sets
JOURNAL Patent: WO 0194629-A 6369 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES
source
1..1721
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Alignment Scores:
Pred. No.: 0.0147 Length: 1721
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 6 Gaps: 0

US-09-606-910E-1 (1-21) x AX335860 (1-1721)

QY 1 AlaHisGlyValThrSerAlaProGlyThrArgProAlaProGlySerThrAlaProPro 20
Db 580 GCCCAGGGTGTACCTCGGCCCGGACACAGCGCGGCCCGGGCTCCACCGCCCCCA 639

QY 21 Ala 21
Db 640 GCC 642

RESULT 14
AX440427
LOCUS AX440427 1721 bp DNA linear PAT 28-JUN-2002
DEFINITION Sequence 280 from Patent WO0190154.
ACCESSION AX440427
VERSION AX440427.1 GI:21665237
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1 Xu, J., Mitcham, J.L., Harlocker, S.L., Dillon, D.C., Secrist, H.,
Lodes, M.J., Algate, P.A., Fling, S.P., Mannion, J., Benson, D.R. and
Carter, D.
TITLE Compositions and methods for the therapy and diagnosis of ovarian
cancer

JOURNAL Patent: WO 0190154-A 280 29-NOV-2001;
CORIXA CORPORATION (US)
FEATURES
source
1..1721
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Alignment Scores:
Pred. No.: 0.0147 Length: 1721
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 6 Gaps: 0

US-09-606-910E-1 (1-21) x AX587588 (1-1721)

QY 1 AlaHisGlyValThrSerAlaProGlyThrArgProAlaProGlySerThrAlaProPro 20
Db 580 GCCCAGGGTGTACCTCGGCCCGGACACAGCGCGGCCCGGGCTCCACCGCCCCCA 639

QY 21 Ala 21
Db 640 GCC 642

RESULT 15
AX587588
LOCUS AX587588 1721 bp DNA linear PAT 10-JAN-2003
DEFINITION Sequence 58 from Patent WO0246467.
ACCESSION AX587588
VERSION AX587588.1 GI:28212336
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 Bertucci, F., Houlgatte, R., Birnbaum, D., Nguyen, C., Viens, P. and
Pert, V.
TITLE Gene expression profiling of primary breast carcinomas using arrays
of candidate genes
JOURNAL Patent: WO 0246467-A 58 13-JUN-2002;
Ipsogen (FR)
FEATURES
source
1..1721
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/misc_feature
1..1721
/note="mucin 1, transmembrane (MUC1) gene."
ORIGIN
Alignment Scores:
Pred. No.: 0.0147 Length: 1721
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 6 Gaps: 0

US-09-606-910E-1 (1-21) x AX587588 (1-1721)

QY 1 AlaHisGlyValThrSerAlaProGlyThrArgProAlaProGlySerThrAlaProPro 20
Db 580 GCCCAGGGTGTACCTCGGCCCGGACACAGCGCGGCCCGGGCTCCACCGCCCCCA 639

QY 21 Ala 21
Db 640 GCC 642
```

Search completed: April 14, 2006, 19:45:51
Job time : 1736.78 secs

This Page Blank (uspto)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 14, 2006, 18:40:06 ; Search time 387.22 Seconds

(without alignments)
361.445 Million cell updates/sec

Title: US-09-606-910E-1

Perfect score: 113

Sequence: 1 AHGVTSAEDTRPAGSTAPPA 21

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4996997 seqs, 332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_p2n.model -DEV=xlh
-Q=/abs/ABSWEB/spool/US09606910/runat_14042006_091618_1498/app.query.fasta.1
-DB=N_Geneseq -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNIT5=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abs08
-USER=US09606910_@CGN_1_1_900_@runat_14042006_091618_1498 -NCPU=6 -ICPU=3
-NO MMAP -NEG SCORES=0 -WAIT -DSBLOK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N_Geneseq_21.*

1: Geneseqn1980s.*
2: Geneseqn1990s.*
3: Geneseqn2000s.*
4: Geneseqn2001as.*
5: Geneseqn2001bs.*
6: Geneseqn2002as.*
7: Geneseqn2002bs.*
8: Geneseqn2003as.*
9: Geneseqn2003bs.*
10: Geneseqn2003cs.*
11: Geneseqn2003ds.*
12: Geneseqn2004as.*
13: Geneseqn2004bs.*
14: Geneseqn2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	113	100.0	120	14	ADX05274 Human muc
C 2	113	100.0	156	10	ADK68635 HSP65-MUC
C 3	113	100.0	162	14	ADK05279 HSP65-MUC
C 4	113	100.0	309	1	AAN90579 pDF9.3 cd

5	113	100.0	328	12	ADIS7670	Adi57670 Human bre
6	113	100.0	525	3	AAD00385	Aad00385 Human muc
7	113	100.0	891	3	AAD00391	Aad00391 Ubiquitin
8	113	100.0	1125	14	ADZ87487	Adz87487 Human muc
9	113	100.0	1194	12	ADIS7712	Adi57712 Human bre
10	113	100.0	1371	3	AD000388	Ad000388 Human muc
11	113	100.0	1378	12	ADIS7693	Adi57693 Human bre
12	113	100.0	1424	12	ADO23180	Ado23180 Antisense
13	113	100.0	1428	6	ABL60159	Abi60159 Human muc
14	113	100.0	1428	12	ADO23125	Ado23125 Human muc
15	113	100.0	1457	12	ADF32627	Adf32627 Plasmid J
16	113	100.0	1527	2	AAV48329	Aav48329 MiniMUC1
17	113	100.0	1548	14	AEA11045	Aea11045 DNA encod
18	113	100.0	1548	14	AEA12738	Aea12738 Wobbled M
19	113	100.0	1548	14	AEBO0289	Aeb00289 DNA encod
20	113	100.0	1572	5	AAS00585	Aas00585 Human muc
21	113	100.0	1614	12	ADK70370	Adk70370 Respirato
22	113	100.0	1630	12	ADIS7708	Adi57708 Human bre
23	113	100.0	1634	12	ADIS7689	Adi57689 Human bre
24	113	100.0	1712	12	ADIS7686	Adi57686 Human bre
25	113	100.0	1713	13	ADW78695	Adw78695 Human muc
26	113	100.0	1721	6	ABS76475	Abs76475 cDNA enco
27	113	100.0	1721	6	ABL68032	Abi68032 Ovary can
28	113	100.0	1721	6	ABK09743	Abk09743 Human ova
29	113	100.0	1721	6	ABV94067	Abv94067 Breast ca
30	113	100.0	1721	8	ADA50566	Ada50566 Mucin 1 (
31	113	100.0	1721	8	ACF12906	Acf12906 Human cer
32	113	100.0	1721	9	AAD56938	Aad56938 Human muc
33	113	100.0	1721	12	ADK65926	Adk65926 Human imm
34	113	100.0	1721	13	ADR24914	Adr24914 Breast ca
35	113	100.0	1737	3	AD000394	Ad000394 Ubiquitin
36	113	100.0	1738	12	ADIS7669	Adi57669 Human bre
37	113	100.0	1755	12	ADIS7673	Adi57673 Human bre
38	113	100.0	1761	14	ADX05272	Adx05272 HSP65-MUC
39	113	100.0	1774	12	ADK43991	Adk43991 Plasmid J
40	113	100.0	1774	12	ADF32625	Adf32625 Plasmid J
41	113	100.0	1799	12	ADO23124	Ado23124 Human muc
42	113	100.0	1800	13	ADR89853	Adr89853 Human muc
43	113	100.0	1803	12	ADIS7699	Adi57699 Human bre
44	113	100.0	1804	6	ABL67539	Abi67539 Thyroid c
45	113	100.0	1804	9	AAD56950	Aad56950 Human muc

ALIGNMENTS

RESULT 1
ADX05274
ID ADX05274 standard; DNA; 120 BP.
XX
AC ADX05274;
XX
DT 21-APR-2005 (first entry)
XX
DE Human mucin MUC1 coding sequence, SEQ ID 3.
XX
KW Fusion protein; Cytostatic; Gene Therapy; mucin; MUC1; carcinoma; gene;
KW ds; chromosome 1.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..120
FT /*tag= a
FT /product= "MUC1 epitope"
FT /partial
FT /note= "No start or stop codon"

US2005031649-A1.

10-FEB-2005.

06-AUG-2003; 2003US-00635211.

XX

```
PR 06-AUG-2003; 2003US-00635211.
XX
XX (YUYU/) YU Y.
PA (LIHH/) LI H.
XX (WANG/) WANG L.
XX
PI Yu Y, Li H, Wang L;
XX
XX WPI; 2005-131970/14.
DR P-PSDB; ADX05275.
XX
XX New recombinant fusion protein comprises Bacillus Calmette Guerin (BCG)
PT heat shock protein 65 and 1-5 copies of an epitope of MUC1, useful for
PT treating or preventing human MUC1 expressing carcinomas.
XX
XX Disclosure; SEQ ID NO 3; 19pp; English.
XX
XX The present invention relates to a novel recombinant fusion protein
CC (HSP65-MUC1-ME, ADX05273) which comprises Bacillus Calmette Guerin (BCG)
CC heat shock protein 65 (HSP65) and an epitope of mucin protein MUC1 (MUC1-
CC ME). The BCG HSP65 is located at the amino terminal portion of the fusion
CC protein and the epitope of MUC1 is located at the carboxy terminal
CC portion of the fusion protein. The MUC1 epitope can generate MUC1
CC specific cytotoxic T lymphocytes. The recombinant fusion protein is
CC useful for treating or preventing human carcinomas, preferably MUC1
CC expressing carcinomas. The present sequence is the coding sequence for a
CC MUC1 epitope, used to produce the fusion protein of the invention. The
CC human MUC1 gene is located on chromosome 1q21.
XX
XX Sequence 120 BP; 11 A; 50 C; 32 G; 27 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 5.1e-05 Length: 120
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 14 Gaps: 0

US-09-606-910E-1 (1-21) x ADX05274 (1-120)

QY 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db 19 GCTACGGGTGTTACCTCTGCTCCGGACACCCGCTCCGGCTCCGGGTCTACCGCTCCGCG 78
QY 21 Ala 21
Db 79 GCT 81

RESULT 2
ADK68635/c
ID ADK68635 standard; cDNA; 156 BP.
XX
XX ADK68635;
XX
XX 06-MAY-2004 (first entry)
XX
XX HSP65-MUC1 antigen CTL epitope related cDNA #3.
XX
XX Human breast cancer; Mycobacterium bovis; heat shock protein 65; BCG;
KW HSP65; cytotoxic T lymphocyte; CTL; MUC1; HSP65-MUC1 antigen CTL epitope;
KW ss.
XX
XX Unidentified.
XX
XX CN1368384-A.
XX
XX 11-SEP-2002.
XX
XX 08-FEB-2001; 2001CN-00102614.
XX
XX 08-FEB-2001; 2001CN-00102614.
XX

PA (DIWE-) DIWEIHUAYU BIO TECHNOLOGY CO LTD BEIJING.
XX
XX Yu Y, Li H;
XX
XX WPI; 2003-854662/80.
XX
XX Genetically engineered vaccine of MUC-1 antigen for human breast cancer.
XX
XX Example 2; Page 4 (Disclosure); 14pp; Chinese.
XX
XX The invention relates to a method of preparation of a genetically
CC engineered vaccine for preventing and treating human breast cancer. The
CC method comprises fusing the coding gene of the Mycobacterium bovis heat
CC shock protein 65 (BCG HSP65) with the cytotoxic T lymphocyte (CTL)
CC epitope gene of MUC1 (antigen cell expressed by human breast cancer
CC cells) and the HSP65-MUC1 antigen CTL epitope is expressed in Escherichia
CC coli cells. This sequence represents DNA used in the method of the
CC invention.
XX
XX Sequence 156 BP; 37 A; 42 C; 58 G; 19 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 6.66e-05 Length: 156
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 10 Gaps: 0

US-09-606-910E-1 (1-21) x ADK68635 (1-156)

QY 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db 114 GCTACGGGTGTTACCTCTGCTCCGGACACCCGCTCCGGGTCTACCGCTCCGCG 55
QY 21 Ala 21
Db 54 GCT 52

RESULT 3
ADX05279/c
ID ADX05279 standard; DNA; 162 BP.
XX
XX ADX05279;
XX
XX 21-APR-2005 (first entry)
XX
XX HSP65-MUC1-ME fusion protein reverse PCR primer, SEQ ID 8.
XX
XX Fusion protein; Cytostatic; Gene Therapy; heat shock protein 65; HSP65;
KW mucin; MUC1; carcinoma; PCR; primer; ss.
XX
XX Mycobacterium bovis BCG.
OS Homo sapiens.
OS Synthetic.
XX
XX US2005031649-A1.
XX
XX 10-FEB-2005.
XX
XX 06-AUG-2003; 2003US-00635211.
XX
XX 06-AUG-2003; 2003US-00635211.
XX
XX (YUYU/) YU Y.
PA (LIHH/) LI H.
XX (WANG/) WANG L.
XX
XX Yu Y, Li H, Wang L;
XX
XX WPI; 2005-131970/14.
XX
XX New recombinant fusion protein comprises Bacillus Calmette Guerin (BCG)
```

PT heat shock protein 65 and 1-5 copies of an epitope of MUC1, useful for
PT treating or preventing human MUC1 expressing carcinomas.

XX Example 1; SEQ ID NO 8; 19pp; English.

CC The present invention relates to a novel recombinant fusion protein
CC (HSP65-MUC1-ME, ADX05273) which comprises Bacillus Calmette Guerin (BCG)
CC heat shock protein 65 (HSP65) and an epitope of mucin protein MUC1 (MUC1-
CC ME). The BCG HSP65 is located at the amino terminal portion of the fusion
CC protein and the epitope of MUC1 is located at the carboxy terminal
CC portion of the fusion protein. The MUC1 epitope can generate MUC1
CC specific cytotoxic T lymphocytes. The recombinant fusion protein is
CC useful for treating or preventing human carcinomas, preferably MUC1
CC expressing carcinomas. The present sequence was used in an example for
CC producing the fusion protein of the invention.

XX Sequence 162 BP; 39 A; 43 C; 58 G; 22 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 6.92e-05 Length: 162
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 14 Gaps: 0

US-09-606-910E-1 (1-21) x ADX05279 (1-162)

QY 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
DB 120 GCTACGGGTGTACCTCTGCTCCGGACACCGCTCCGGCTCCGGGTTCACCGCTCCGGCG 61

QY 21 Ala 21
DB 60 GCT 58

RESULT 4

AAN90579/c
ID AAN90579 standard; cDNA; 309 BP.

XX AAN90579;

AC 27-AUG-2003 (revised)

DT 25-MAR-2003 (revised)

DT 04-DEC-1989 (first entry)

XX pDF9.3 cDNA insert.

DE pDF9.3; human DF3 breast carcinoma-associated antigen epitope.

KW Human MCF-7 breast carcinoma cells; 'ATCC HTB22'.

OS WO8907107-A.

FN 10-AUG-1989.

XX 29-JAN-1988; 88US-00149831.

XX 29-JAN-1988; 88US-00149831.

XX (DANA-) DANA-FARBER CANCER.

XX Kufe DW;

XX WPI; 1989-248989/34.

DR P-PSDB; AAP91045, AAP91053, AAP91054, AAP90146.

XX Recombinant polypeptide(s) - contains DF 3 breast carcinoma antigen

PT epitope and useful as assay reagents, and encoding DNA sequences.

XX Claim 1; Fig 4; 31pp; English.

XX The sequence encodes a carbohydrate-free polypeptide contg. DF3 breast

CC carcinoma antigen epitope. Useful as a competitive binding assay reagent
CC and improves diagnosis. The cDNA consists of nearly identical 60 BP
CC tandem repeats which are 85% GC-rich. See also AAP91045, AAP91046,
CC AAP91053, AAP91054. (Updated on 25-MAR-2003 to correct PA field.)
CC (Updated on 27-AUG-2003 to correct OS field.)

XX Sequence 309 BP; 21 A; 92 C; 160 G; 36 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 0.000133 Length: 309
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 1 Gaps: 0

US-09-606-910E-1 (1-21) x AAN90579 (1-309)

QY 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
DB 308 GCCACGGGTGTACCTCGGCCCGGACACCGAGCGCGGCTCCACCGCCCCCA 249

QY 21 Ala 21
DB 248 GCC 246

RESULT 5

ADI57670

ID ADI57670 standard; cDNA; 328 BP.

XX ADI57670;

XX 22-APR-2004 (first entry)

XX Human breast specific nucleic acid (BSNA) #41.

KW Human; breast specific nucleic acid; BSNA; gene; ss; metastasis;

KW breast cancer; cytostatic.

OS Homo sapiens.

XX WO2003106648-A2.

XX 24-DEC-2003.

XX 16-JUN-2003; 2003WO-US018934.

XX 14-JUN-2002; 2002US-0389327P.

XX (DIAD-) DIADEXUS INC.

XX Salceda S, Macina RA, Turner LR, Sun Y, Liu C;

XX WPI; 2004-082185/08.

XX P-PSDB; ADI57744.

XX Novel isolated polypeptide comprising breast specific protein sequences,
XX useful for diagnosing or monitoring presence and metastases of breast
XX cancer in patient.

XX Claim 1; SEQ ID NO 41; 370pp; English.

XX The invention relates to human breast specific nucleic acids (BSNA) and
XX the breast specific proteins (BSP) they encode. The nucleic acids are
XX useful for determining the presence of a BSNA in a sample which involves
XX contacting the sample with a BSNA under conditions in which the BSNA will
XX selectively hybridise to a BSNA in the sample, and detecting the
XX hybridisation. The nucleic acids are useful for determining the presence
XX of a BSP in a sample which involves contacting the sample with suitable
XX reagent under conditions in which the reagent will selectively interact
XX with the BSP, and detecting the interaction of the reagent with a BSP in
XX the sample. The nucleic acids and proteins are useful for diagnosing or
XX monitoring the presence and metastases of breast cancer in a patient.

CC which involves determining an amount of nucleic acid or protein and
 CC comparing the determined amount of nucleic acid or protein in the sample
 CC of the patient to the amount of a breast specific marker in a normal
 CC control, where a difference in the determined amount in the sample
 CC compared to the amount in the control is associated with the presence of
 CC breast cancer. The sequences are useful for treating a patient with
 CC breast cancer, involving administering a composition consisting of a BSN
 CC or a BSP to a patient, where the administration induces an immune
 CC response against the breast cancer cell expressing the BSN or BSP. This
 CC sequence represents a human BSN of the invention.

XX
 SQ Sequence 328 BP; 47 A; 124 C; 95 G; 61 T; 0 U; 1 Other;

Alignment Scores:
 Pred. No.: 0.000142 Length: 328
 Score: 113.00 Matches: 21
 Percent Similarity: 100.0% Conservative: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 100.0% Indels: 0
 DB: 12 Gaps: 0

US-09-606-910E-1 (1-21) x ADI57670 (1-328)

QY 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
 DB 113 GCCACGGGTGTCACCTCGGCCCGCCGACACAGCGCGCCGGGTCCACCGCCCCCA 172

QY 21 Ala 21
 DB 173 GCC 175

RESULT 6

AAD00385
 ID AAD00385 standard; DNA; 525 BP.

XX
 AC AAD00385;

XX 29-AUG-2000 (first entry)

XX Human Mucin 1 (MUC-1) protein fragment encoding DNA #2.

XX Human; Mucin 1; MUC-1; tumour; pMRS30 expression vector; anti-tumour;
 KW therapy; immune response; cytostatic; vaccine; ds.

XX Homo sapiens.

XX Key Location/Qualifiers
 FH CDS 1..525

FT /*tag= a
 FT /product= "MUC-1 protein fragment"

XX WO200025827-A2.

XX 11-MAY-2000.

XX 18-OCT-1999; 99WO-EP007874.

XX 30-OCT-1998; 98IT-MI02330.

XX (MENA) MENARINI RICERCHE SPA.

XX Parente D, Di Massimo AM, De Santis R;

XX WPI; 2000-365410/31.

XX P-PSDB; AAY71021.

XX Composition containing one or more DNA molecules encoding fragments of a
 PT Mucin 1 (MUC-1) protein overexpressed in tumor cells, useful in anti-
 PT tumor therapy.

XX Claim 16; Fig 2; 56pp; English.

XX The present sequence is a DNA encoding a fragment of human Mucin 1 (MUC-

CC 1) antigenic protein which is overexpressed in tumour cells. The sequence
 CC was obtained from BT20 tumour cells by reverse transcriptase-PCR and
 CC corresponds to nucleotides 205-720 of the EMBL sequence J05581 with a
 CC start codon and two stop codons. The present sequence is cloned into a
 CC pMRS30 expression vector and used in pharmaceutical composition e.g.
 CC vaccine for inducing an antigen-specific anti-tumour immune response.
 CC Composition containing this DNA molecule is useful in anti-tumour therapy
 CC of patients affected with tumours characterised by high MUC-1 expression

XX Sequence 525 BP; 111 A; 199 C; 128 G; 87 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 0.000229 Length: 525
 Score: 113.00 Matches: 21
 Percent Similarity: 100.0% Conservative: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 100.0% Indels: 0
 DB: 3 Gaps: 0

US-09-606-910E-1 (1-21) x AAD00385 (1-525)

QY 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
 DB 274 GCACACGGGTGTACCTCGGCTCCGATACACAGCGCGCCAGGTAGTACCGCCCTCT 333

QY 21 Ala 21

DB 334 GCC 336

RESULT 7

AAD00391

ID AAD00391 standard; DNA; 891 BP.

XX
 AC AAD00391;

XX 15-SEP-2003 (revised)

XX 29-AUG-2000 (first entry)

XX Ubiquitin-E. coli LacI-human Mucin 1 fusion protein encoding DNA #2.

XX Ubiquitin; LacI; beta-galactosidase; fusion protein; human; Mucin 1;
 KW MUC-1; tumour; pMRS30 expression vector; anti-tumour; therapy;
 KW immune response; cytostatic; vaccine; ds.

XX Homo sapiens.

XX Escherichia coli.
 OS Chimeric.

XX Key Location/Qualifiers
 FH CDS 1..891

FT /*tag= a
 FT /product= "Ubiquitin-LacI-MUC-1 fusion protein"

FT misc_feature 1..369

FT /*tag= b
 FT /label= UBILacI DNA

FT /*note= "Includes ubiquitin-E. coli LacI fusion DNA"

FT misc_feature 370..891

FT /*tag= c

FT /*note= "Human MUC-1 partial DNA that corresponds to
 FT nucleotides 205-720 of the EMBL sequence J05581 with two
 FT stop codons"

XX WO200025827-A2.

XX 11-MAY-2000.

XX 18-OCT-1999; 99WO-EP007874.

XX 30-OCT-1998; 98IT-MI02330.

XX (MENA) MENARINI RICERCHE SPA.

XX Parente D, Di Massimo AM, De Santis R;


```
XX WPI; 2000-365410/31.
DR DR
XX P-PSDB; AA71027.
XX Composition containing one or more DNA molecules encoding fragments of a
PT Mucin 1 (MUC-1) protein overexpressed in tumor cells, useful in anti-
PT tumor therapy.
XX Claim 18; Fig 8; 56pp; English.
XX The present sequence is a DNA encoding a fusion protein consisting of
CC human Mucin 1 (MUC-1) fragment fused to UBI-LACI sequence at the N-
CC terminus. The UBI-LACI sequence consists of ubiquitin from MCF7 cell line
CC and a portion of E. coli beta-galactosidase (LacI). MUC-1 is an antigenic
CC protein overexpressed in tumour cells. The present sequence is cloned
CC into a pMR330 expression vector and used in pharmaceutical composition
CC e.g. vaccine for inducing an antigen-specific anti-tumour immune
CC response. Composition containing this DNA molecule is useful in anti-
CC tumour therapy of patients affected with tumours characterised by high
CC MUC-1 expression. (Updated on 15-SEP-2003 to standardise OS field)
XX SQ Sequence 891 BP; 203 A; 307 C; 225 G; 156 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 0.000392 Length: 891
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 3 Gaps: 0
US-09-606-910E-1 (1-21) x AAD00391 (1-891)
QY 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
DB 640 GCACAGCGGTGTACTCGGATCCGGATACAGCCGCGCCCGAGTAGTACCGCCCTCT 699
QY 21 Ala 21
DB 700 GCC 702
RESULT 8
ADZ87487
ID ADZ87487 standard; DNA; 1125 BP.
XX AC ADZ87487;
XX 14-JUL-2005 (first entry)
XX Human MUC1-EC protein-encoding gene SeqID14.
XX protein engineering; immunoglobulin; cytostatic; cancer; gene; ds.
XX Homo sapiens.
XX WO2005042573-A1.
XX 12-MAY-2005.
XX 21-OCT-2004; 2004WO-US034680.
XX 24-OCT-2003; 2003US-0514198P.
XX 12-NOV-2003; 2003US-0519822P.
XX (DAND ) DANA FARBER CANCER INST INC.
XX (ILEX-) ILEX PROD INC.
XX Kharbanda S, Kufe DW;
XX WPI; 2005-346855/35.
XX P-PSDB; ADZ87486.
XX MUC1 ligand trap comprising a MUC1 extracellular polypeptide and an
PT immunoglobulin FC polypeptide or an albumin polypeptide, useful for
PT treating cancer.
XX Disclosure; SEQ ID NO 14; 82pp; English.
XX This invention relates to a novel MUC1 chimeric protein which comprises a
CC first polypeptide sequence and a second polypeptide sequence, where the
CC first polypeptide sequence is a MUC1 extracellular (MUC1-EC) polypeptide,
CC and the second polypeptide sequence is a human immunoglobulin FC
CC polypeptide or a human albumin polypeptide. The invention may be useful
CC for the development of compounds with a cytostatic activity acting as
CC MUC1 antagonists or CD227 antagonists. The MUC1 chimeric protein is
CC useful in preparing a composition for treating cancer. The present
CC sequence is that of a gene which was used during the development of the
CC novel MUC1 chimeric protein of the invention.
XX SQ Sequence 1125 BP; 240 A; 402 C; 246 G; 237 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 0.000497 Length: 1125
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 14 Gaps: 0
US-09-606-910E-1 (1-21) x ADZ87487 (1-1125)
QY 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
DB 403 GCCACAGGTGTACCTCGGCCCGGACACAGCGCGCGGCTCCACCGCCCCCA 462
QY 21 Ala 21
DB 463 GCC 465
RESULT 9
AD157712
ID AD157712 standard; cDNA; 1194 BP.
XX AC AD157712;
XX 22-APR-2004 (first entry)
XX Human breast specific nucleic acid (BSNA) #83.
XX Human; breast specific nucleic acid; BSNA; gene; ss; metastasis;
XX breast cancer; cytostatic.
XX Homo sapiens.
XX WO2003106648-A2.
XX 24-DEC-2003.
XX 16-JUN-2003; 2003WO-US018934.
XX 14-JUN-2002; 2002US-0389327P.
XX (DIAD-) DIADEXUS INC.
XX Salceda S, Macina RA, Turner LR, Sun Y, Liu C;
XX WPI; 2004-082185/08.
XX P-PSDB; AD157782.
XX Novel isolated polypeptide comprising breast specific protein sequences,
PT useful for diagnosing or monitoring presence and metastases of breast
PT cancer in patient.
XX Claim 1; SEQ ID NO 83; 370pp; English.
XX The invention relates to human breast specific nucleic acids (BSNA) and
CC
```

CC the breast specific proteins (BSP) they encode. The nucleic acids are
 CC useful for determining the presence of a BSNA in a sample which involves
 CC contacting the sample with a BSNA under conditions in which the BSNA will
 CC selectively hybridise to a BSNA in the sample, and detecting the
 CC hybridisation. The nucleic acids are useful for determining the presence
 CC of a BSP in a sample which involves contacting the sample with suitable
 CC reagent under conditions in which the reagent will selectively interact
 CC with the BSP, and detecting the interaction of the reagent with a BSP in
 CC the sample. The nucleic acids and proteins are useful for diagnosing or
 CC monitoring the presence and metacases of breast cancer in a patient,
 CC which involves determining an amount of nucleic acid or protein and
 CC comparing the determined amount of nucleic acid or protein in the sample
 CC of the patient to the amount of a breast specific marker in a normal
 CC control, where a difference in the determined amount in the sample
 CC compared to the amount in the control is associated with the presence of
 CC breast cancer. The sequences are useful for treating a patient with
 CC breast cancer, involving administering a composition consisting of a BSNA
 CC or a BSP to a patient, where the administration induces an immune
 CC response against the breast cancer cell expressing the BSNA or BSP. This
 CC sequence represents a human BSNA of the invention.

XX Sequence 1194 BP; 243 A; 438 C; 275 G; 238 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 0.000528 Length: 1194
 Score: 113.00 Matches: 21
 Percent Similarity: 100.0% Conservative: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 100.0% Indels: 0
 DB: 12 Gaps: 0

US-09-606-910E-1 (1-21) x ADI57712 (1-1194)

QY 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
 DB 614 GCCACGGTGTCACTCGCGCCCGGACACACAGCGCGGCCCGGCTCCACCGCCCCCA 673

QY 21 Ala 21

DB 674 GCC 676

RESULT 10

AAD00388

ID AAD00388 standard; DNA; 1371 BP.

XX AC AAD00388;

XX 29-AUG-2000 (first entry)

XX Human Mucin 1 (MUC-1) protein fragment encoding DNA #5.

XX Human; Mucin 1; tumour; pMRS30 expression vector; anti-tumour;
 XX therapy; immune response; cytostatic; vaccine; ds.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 1..1371

XX /*tag= a

XX /product= "MUC-1 protein fragment"

XX WO200025827-A2.

XX 11-MAY-2000.

XX 18-OCT-1999; 99WO-EP007874.

XX 30-OCT-1998; 98IT-MI002330.

XX (MENA) MENARINI RICERCH SPA.

XX Parente D, Di Massimo AM, De Santis R;

XX

DR WPI; 2000-365410/31.
 DR P-PSDB; AAY71024.

XX Composition containing one or more DNA molecules encoding fragments of a
 PT Mucin 1 (MUC-1) protein overexpressed in tumor cells, useful in anti-
 PT tumor therapy.

PS Claim 16; Fig 5; 56pp; English.

XX The present sequence is a DNA encoding a fragment of human Mucin 1 (MUC-
 CC 1) antigenic protein which is overexpressed in tumour cells. The sequence
 CC was obtained by PCR from plasmids pMRS166, pMRS167, pMRS168 and pMRS169
 CC which contain MUC-1 DNA from BT20 tumour cells. It corresponds to
 CC nucleotides 136-1497 of the EMBL sequence J05581 with a start codon and
 CC two stop codons. The present sequence is cloned into a pMRS30 expression
 CC vector and used in pharmaceutical composition e.g. vaccine for inducing
 CC an antigen-specific anti-tumour immune response. Composition containing
 CC this DNA molecule is useful in anti-tumour therapy of patients affected
 CC with tumours characterised by high MUC-1 expression

XX Sequence 1371 BP; 300 A; 450 C; 321 G; 300 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 0.000607 Length: 1371
 Score: 113.00 Matches: 21
 Percent Similarity: 100.0% Conservative: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 100.0% Indels: 0
 DB: 3 Gaps: 0

US-09-606-910E-1 (1-21) x AAD00388 (1-1371)

QY 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
 DB 343 GCACACGGTGTACCTCGGCTCCGATACACAGCGCGGCCCGGCTCTCTCTCTCTCTCT 402

QY 21 Ala 21

DB 403 GCC 405

RESULT 11

ADI57693

ID ADI57693 standard; cDNA; 1378 BP.

XX AC ADI57693;

XX 22-APR-2004 (first entry)

XX Human breast specific nucleic acid (BSNA) #64.

XX Human; breast specific nucleic acid; BSNA; gene; ss; metastasis;
 XX breast cancer; cytostatic.

XX Homo sapiens.

XX WO2003106648-A2.

XX 24-DEC-2003.

XX 16-JUN-2003; 2003WO-US018934.

XX 14-JUN-2002; 2002US-0389327P.

XX (DIAD-) DIADEXUS INC.

XX Salceda S, Macina RA, Turner LR, Sun Y, Liu C;

XX WPI; 2004-082185/08.

XX P-PSDB; ADI57765.

XX Novel isolated polypeptide comprising breast specific protein sequences,
 PT useful for diagnosing or monitoring presence and metastases of breast
 PT cancer in patient.

```
XX Claim 1; SEQ ID NO 64; 370pp; English.
XX
XX The invention relates to human breast specific nucleic acids (BSNA) and
XX the breast specific proteins (BSP) they encode. The nucleic acids are
XX useful for determining the presence of a BSNA in a sample which involves
XX contacting the sample with a BSNA under conditions in which the BSNA will
XX selectively hybridise to a BSNA in the sample, and detecting the
XX hybridisation. The nucleic acids are useful for determining the presence
XX of a BSP in a sample which involves contacting the sample with suitable
XX reagent under conditions in which the reagent will selectively interact
XX with the BSP, and detecting the interaction of the reagent with a BSP in
XX the sample. The nucleic acids and proteins are useful for diagnosing or
XX monitoring the presence and metastases of breast cancer in a patient,
XX which involves determining an amount of nucleic acid or protein and
XX comparing the determined amount of nucleic acid or protein in the sample
XX of the patient to the amount of a breast specific marker in a normal
XX control, where a difference in the determined amount in the sample
XX compared to the amount in the control is associated with the presence of
XX breast cancer. The sequences are useful for treating a patient with
XX breast cancer, involving administering a composition consisting of a BSNA
XX or a BSP to a patient, where the administration induces an immune
XX response against the breast cancer cell expressing the BSNA or BSP. This
XX sequence represents a human BSNA of the invention.
XX
XX SQ Sequence 1378 BP; 295 A; 489 C; 316 G; 278 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 0.000611 Length: 1378
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 12 Gaps: 0

US-09-606-910E-1 (1-21) x ADI57693 (1-1378)
QY 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db 614 GCCCAGGGTGTCACTTCGGCCCGGACACAGCGCCGGCCCGGGCTCCACCGCCCCCA 673
QY 21 Ala 21
Db 674 GCC 676

RESULT 12
ADO23180/c
ID ADO23180 standard; RNA; 1424 BP.
XX
XX ADO23180;
XX
XX 12-AUG-2004 (first entry)
XX
XX Antisense human MUC1 mucin glycoprotein RNA (coding sequence) SeqID 75.
XX human; MUC1; mucin glycoprotein; cancer; chemotherapeutic; MUC1/BCD; ss;
XX antisense.
XX
XX Homo sapiens.
XX
XX WO2004044160-A2.
XX
XX 27-MAY-2004.
XX
XX 12-NOV-2003; 2003WO-US035848.
XX
XX 13-NOV-2002; 2002US-00293391.
XX
XX 29-MAY-2003; 2003US-00447839.
XX
XX (DAND ) DANA FARBER CANCER INST INC.
XX
XX (ILEX-) ILEX PROD INC.
XX
XX Kufe DW, Kharbanda S, Weitman SD;

XX WPI; 2004-420304/39.
XX
XX Double-stranded RNA complex useful for inhibiting proliferation of cancer
XX cell expressing MUC1 mucin glycoprotein, comprises first and second RNA
XX sequences.
XX
XX Disclosure; SEQ ID NO 75; 112pp; English.
XX
XX This invention relates to novel modulators of the human MUC1 mucin
XX glycoprotein for use in cancer therapeutics, where MUC1 is a protein that
XX acts to inhibit the apoptotic response to genotoxic stress caused by
XX chemotherapeutic agents. In particular, it refers to modulators of the
XX MUC1 extracellular domain (MUC1/BCD). The method refers to using double-
XX stranded RNA complexes as MUC1 interference RNA compositions such that
XX MUC1 expression is inhibited, which in turn inhibits cancer cell
XX proliferation. The present invention describes screening assays to
XX identify compounds that inhibit the binding of various MUC1 ligands such
XX as neuregulin 2, as well as agonists, antagonists and antibodies thereof.
XX Furthermore, it provides MUC1 antisense, small interfering RNA (siRNA)
XX and small molecules in combination with chemotherapeutic agents that are
XX useful in the field of cancer therapy. This polynucleotide sequence is
XX the antisense human MUC1 RNA of the invention.
XX
XX SQ Sequence 1424 BP; 309 A; 325 C; 495 G; 0 T; 295 U; 0 Other;

Alignment Scores:
Pred. No.: 0.000631 Length: 1424
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 12 Gaps: 0

US-09-606-910E-1 (1-21) x ADO23180 (1-1424)
QY 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db 1023 GCCCAGGGTGTCACTTCGGCCCGGACACAGCGCCGGCCCGGGCTCCACCGCCCCCA 964
QY 21 Ala 21
Db 963 GCC 961

RESULT 13
ABL60159
ID ABL60159 standard; cDNA; 1428 BP.
XX
XX ABL60159;
XX
XX 22-JUL-2002 (first entry)
XX
XX Human MUC1 encoding cDNA SEQ ID NO 2.
XX
XX Human; mucin 1; MUC1; transmembrane protein; SNP; cancer; cytostatic;
XX single nucleotide polymorphism; haplotyping; genotyping; drug;
XX antiinflammatory; gene; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 1..1428
XX FT /*tag= a
XX FT /product= "MUC1"
XX FT replace(1009,A)
XX FT /*tag= b
XX FT variation
XX
XX /standard_name= "Single nucleotide polymorphism"
XX /note= "SNP allelic variation results in Val substituted
XX by Met at position 337 of the MUC1 protein (ABB77476)"

XX WO200226765-A2.
XX
XX 04-APR-2002.
XX
XX PD
```

```

XX 25-SEP-2001; 2001WO-US030151.
PF
XX
XX 28-SEP-2000; 2000US-0236113P.
PR
XX
XX (GENA-) GENAISSANCE PHARM INC.
XX
XX Chew A, Koshiy B;
PI
XX WPI; 2002-405042/43.
DR
DR P-PSDB; ABE77476.
XX
XX New genetic variants of mucin 1, Transmembrane gene, useful in studying
PT expression and function of protein encoded by the gene and for screening
PT drugs to treat diseases e.g. cancer.
XX
XX Claim 23; Fig 2; 75pp; English.
PS
XX The invention relates to a polynucleotide (ABL60159, ABL60159) encoding
CC mucin 1/MUC1 (ABE77476), Transmembrane isogene. The invention describes
CC novel genetic variants of the MUC1 gene. The invention is useful for
CC haplotyping/genotyping the MUC1 gene in an individual and identifying an
CC association between a trait and at least one of the haplotypes or
CC haplotype pairs of MUC1 gene. MUC1 is useful for studying the expression
CC and function of MUC1 and expressing MUC1 protein for use in screening for
CC candidate drugs to treat diseases related to MUC1 activity and in
CC studying the effect of the variation on the biological activity of MUC1
CC as well as on the binding affinity of candidate drugs targeting MUC1 for
CC the treatment of e.g. cancer. MUC1 is further used by the pharmaceutical
CC research scientist to validate MUC1 as a candidate target for and in
CC design of clinical trials of candidate drugs for, treating a specific
CC condition or disease predicted to be associated with MUC1 activity.
CC MUC1 antibodies are useful in a variety of diagnostic and prognostic
CC formats and therapeutic methods
XX
XX Sequence 1428 BP; 296 A; 496 C; 327 G; 309 T; 0 U; 0 Other;
SQ
XX
XX Alignment Scores:
Pred. No.: 0.000633 Length: 1428
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 6 Gaps: 0
XX
XX US-09-606-910E-1 (1-21) x ABL60159 (1-1428)
QY 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db 403 GCCCAGGGTGTCACTCGGCCCCGGACACACGCGCGCGGCTCCACCGCCCCCA 462
QY 21 Ala 21
Db 463 GCC 465
XX
XX RESULT 14
AD023125
ID ADO23125 standard; RNA; 1428 BP.
XX
XX ADO23125;
AC
XX
XX 12-AUG-2004 (first entry)
DT
XX
XX Human MUC1 mucin glycoprotein RNA (coding sequence) SeqID 20.
DE
XX human; MUC1; mucin glycoprotein; cancer; chemotherapeutic; MUC1/ECOD; ss.
KW
XX Homo sapiens.
OS
XX WO2004044160-A2.
PN
XX 27-MAY-2004.
PD
XX
XX

```

```

PF 12-NOV-2003; 2003WO-US035848.
XX
XX 13-NOV-2002; 2002US-00293391.
PR
XX 29-MAY-2003; 2003US-0047839.
XX
XX (DAND ) DANA FARBER CANCER INST INC.
PA
XX (ILEX-) ILEX PROD INC.
PA
XX Kufe DW, Kharbanda S, Weitman SD;
PI
XX WPI; 2004-420304/39.
DR
XX
XX Double-stranded RNA complex useful for inhibiting proliferation of cancer
PT cell expressing MUC1 mucin glycoprotein, comprises first and second RNA
PT sequences.
XX
XX Claim 2; SEQ ID NO 20; 112pp; English.
PS
XX This invention relates to novel modulators of the human MUC1 mucin
CC glycoprotein for use in cancer therapeutics, where MUC1 is a protein that
CC acts to inhibit the apoptotic response to genotoxic stress caused by
CC chemotherapeutic agents. In particular, it refers to modulators of the
CC MUC1 extracellular domain (MUC1/ECOD). The method refers to using double-
CC stranded RNA complexes as MUC1 interference RNA compositions such that
CC MUC1 expression is inhibited, which in turn inhibits cancer cell
CC proliferation. The present invention describes screening assays to
CC identify compounds that inhibit the binding of various MUC1 ligands such
CC as neuregulin 2, as well as agonists, antagonists and antibodies thereof.
CC Furthermore, it provides MUC1 antisense, small interfering RNA (siRNA)
CC and small molecules in combination with chemotherapeutic agents that are
CC useful in the field of cancer therapy. This polynucleotide sequence is
CC the human MUC1 RNA of the invention.
XX
XX Sequence 1428 BP; 296 A; 496 C; 327 G; 0 T; 309 U; 0 Other;
SQ
XX
XX Alignment Scores:
Pred. No.: 0.000633 Length: 1428
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 12 Gaps: 0
XX
XX US-09-606-910E-1 (1-21) x ADO23125 (1-1428)
QY 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db 403 GCCCAGGGTGTCACTCGGCCCCGGACACACGCGCGCGGCTCCACCGCCCCCA 462
QY 21 Ala 21
Db 463 GCC 465
XX
XX RESULT 15
ADP32627
ID ADP32627 standard; DNA; 1457 BP.
XX
XX AC ADP32627;
AC
XX
XX 26-FEB-2004 (first entry)
DT
XX
XX Plasmid JNW358 MUC-1 nucleotide sequence.
DE
XX
XX MUC-1 antigen; immune response; MUC-1; variable number of tandem repeat;
KW VNTR; repeat unit; tumour; metastasis; cytostatic; vaccine; gene therapy;
KW gene; ds.
XX
XX Synthetic.
OS
XX WO2003100060-A2.
PN
XX 04-DEC-2003.
PD
XX
XX

```

PF 23-MAY-2003; 2003WO-EP005594.
XX
PR 24-MAY-2002; 2002GB-00012046.
XX
XX (GLAX) GLAXO GROUP LTD.
XX
PI Burden N, Ellis JH, Hamblin PA;
XX
XX
DR WPI; 2004-042811/04.
XX
PT New nucleic acid molecule encoding a MUC-1 antigen, useful for preparing
PT a composition for treating or preventing tumors or metastases.
XX
XX
PS Example; Fig 3; 66pp; English.
XX
CC The present invention describes a nucleic acid molecule which encodes a
CC MUC-1 antigen. The nucleic acid is capable of raising an immune response
CC in vivo, has reduced susceptibility to recombination than full-length MUC
CC -1 and comprises between 1 and 15 variable number of tandem repeats
CC (VNTR) perfect repeat units. Also described: (1) a plasmid comprising the
CC DNA molecule; (2) a protein encoded by the nucleic acid; (3) a
CC pharmaceutical composition comprising the nucleic acid, plasmid or
CC protein and an excipient, diluent or carrier; and (4) a method of
CC treating or preventing tumors or metastases. A MUC1 antigen has
CC cytostatic activity, and can be used in vaccines, and in gene therapy.
CC The nucleic acid is useful for preparing a composition for treating or
CC preventing tumors or metastases. The present sequence is used in the
CC exemplification of the present invention.
XX
SQ Sequence 1457 BP; 302 A; 506 C; 334 G; 315 T; 0 U; 0 Other;
XX
Alignment Scores:
Pred. No.: 0.000646 Length: 1457
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 12 Gaps: 0
US-09-606-910E-1 (1-21) x ADF32627 (1-1457)
OY 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db 421 GCCCAGGTTGTCACTCGGCCCGGACACAGGCGGGCCCGGGCTCCACCGCCCCCA 480
OY 21 Ala 21
Db 481 GCC 483
Search completed: April 14, 2006, 19:06:33
Job time : 389.22 secs

This page Blank (uspto)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 14, 2006, 18:49:30 ; Search time 3304.17 Seconds
(without alignments)
297.360 Million cell updates/sec

Title: US-09-606-910E-1
Perfect score: 113
Sequence: 1 AHGVTSPADTRPAGSTAPPA 21

Scoring table: BLOSUM62 Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-Q=/abs/ABSSWEB_spool/SUFFIX-p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-DB=EST -QFMT=fastcap -SUFFIX=p2n.rst -TRANS=human40.cdi -LIST=45
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -ALIGN=15 -MODE=LOCAL
-DOALIGN=200 -THRM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abs05p
-OUTFMT=pt0 -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abs05p
-USER=US09606910 @CGN 1 1 10172 @runat 14042006 091621 1556 -NCPU=6 -ICPU=3
-NO_MMAP -NEG_SCORES=0 -WAIT -DSFBLCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

```
EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_hic:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_ges1:*
10: gb_ges2:*
11: gb_ges3:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c	1	113	100.0	330 1	AI925867 wo20d04.x
	2	113	100.0	604 3	BM791359 K-EST0071
	3	113	100.0	754 8	DR422781 nav17d09
	4	113	100.0	877 5	BUS42454 AGENCOURT
	5	113	100.0	959 6	CA489836 AGENCOURT
	6	113	100.0	1113 5	BUI48487 AGENCOURT
	7	113	100.0	1130 5	BUS422996

8	113	100.0	1234	5	BQ936898	BQ936898	AGENCOURT
9	113	100.0	1262	5	BQ935496	BQ935496	AGENCOURT
10	113	100.0	1343	5	BQ920055	BQ920055	AGENCOURT
11	113	100.0	1349	5	BUI52566	BUI52566	AGENCOURT
12	113	100.0	1420	5	BUS42790	BUS42790	AGENCOURT
13	113	100.0	1531	5	BUS43309	BUS43309	AGENCOURT
14	105	92.9	1268	5	BQ943554	BQ943554	AGENCOURT
15	104	92.0	1334	5	BQ943809	BQ943809	AGENCOURT
16	94	83.2	472	3	BM759495	BM759495	K-EST0039
17	94	83.2	475	6	CB120860	CB120860	K-EST0168
18	94	83.2	669	6	CB122585	CB122585	K-EST0170
19	94	83.2	981	2	BG774910	BG774910	602649832
20	94	83.2	1536	5	BQ923149	BQ923149	AGENCOURT
21	90	79.6	166	8	T27692	EST12384	HU
22	85	75.2	619	2	BI260921	BI260921	602970962
23	79	69.9	1678	2	BG775565	BG775565	602650481
24	76	67.3	690	1	AL543598	AL543598	ALS43598
25	76	67.3	1241	4	CR596859	full-length	CR596859
26	67	59.3	1187	3	BQ219669	BQ219669	AGENCOURT
27	67	59.3	1366	3	BM046583	BM046583	603626657
C	28	65	57.5	172	2	BI015399	PM3-RT020
	29	65	57.5	940	2	BE901435	601674774
C	30	64.5	57.1	333	10	CL974028	OB1PCC042
	31	64	56.6	154	2	BF463559	UI-M-CGDP
C	32	64	56.6	344	1	AW494383	UI-M-BH3-
	33	63.5	56.2	533	3	BI795528	H024P07 E
C	34	63	55.8	472	10	CL804119	OR_CBA001
	35	63	55.8	500	1	AV591193	AV591193
C	36	63	55.8	548	2	BE706360	RCL1-H025
	37	63	55.8	708	3	BM582005	170006872
C	38	63	55.8	722	7	CR993345	CR993345
	39	63	55.8	786	2	BG718241	602696220
C	40	62	54.9	336	5	BY783919	BY783919
	41	62	54.9	435	7	CK061420	55059381C
C	42	62	54.9	549	2	BF929925	MK2-NT013
	43	62	54.9	622	7	CV014984	201-E05X
C	44	62	54.9	656	1	AV609791	AV609791
	45	62	54.9	831	10	CZ356841	ZMMBF0096

ALIGNMENTS

RESULT 1

AI925867/c
LOCUS wo20d04.x1 NCI_CGAP Pan1 Homo sapiens cDNA clone IMAGE:2455879 3',
DEFINITION similar to contains element MSRI MSRI repetitive element ;, mRNA
sequence.

ACCESSION AI925867
VERSION AI925867.1 GI:5661831

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.

REFERENCE 1 (bases 1 to 330)

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

JOURNAL Tumor Gene Index

COMMENT Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgaps-r@mail.nih.gov

Life Technologies catalog #: 11548-013

DNA Sequencing by: Washington University Genome Sequencing Center
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 578 Std Error: 0.00

Seq primer: -40UP from Gibco

High quality sequence stop: 305.

Location/Qualifiers

source

1. .330

/organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2455879"
 /tissue_type="adenocarcinoma"
 /lab_host="DH10B"
 /clone_lib="NCI_CGAP_Pan1"
 /note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: SalI;
 Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.72 kb. Life Technologies catalog #: 11548-013"

ORIGIN

Alignment Scores:
 Pred. No.: 0.00121 Length: 330
 Score: 113.00 Matches: 21
 Percent Similarity: 100.0% Conservative: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 100.0% Indels: 0
 DB: 1 Gaps: 0

US-09-606-910E-1 (1-21) x AI925867 (1-330)

Qy 1 AlahisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
 |||||
 Db 298 GCCCAGCGTGTACCTCGGCCCGACACCGCGCGCGCGCTCCACCGCCCCCA 239
 Qy 21 Ala 21
 ||||
 Db 238 GCC 236

RESULT 2
 BM791359
 LOCUS BM791359 604 bp mRNA linear EST 05-MAR-2002
 DEFINITION K-EST0071342 S21SNU520 Homo sapiens cDNA clone S21SNU520-14-A06 5',
 mRNA sequence.

ACCESSION BM791359
 VERSION BM791359.1 GI:19139591
 KEYWORDS EST.

SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Hominidae; Homo.

REFERENCE 1 (bases 1 to 604)
 AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
 Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
 Kim,Y.S.

TITLE 21C Frontier Korean EST Project 2001
 JOURNAL Unpublished (2002)
 COMMENT Genome Research Center
 Korea Research Institute of Bioscience & Biotechnology
 52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
 Tel: +82-42-860-4470
 Fax: +82-42-860-4409
 Email: yongsung@mail.kribb.re.kr
 Plate: 14 row: A column: 06
 High quality sequence stop: 604.
 Location/Qualifiers

FEATURES

source
 1..604
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="S21SNU520-14-A06"
 /sex="F"
 /tissue_type="Stomach"
 /cell_type="Floating aggregates"
 /cell_line="SNU-520"
 /lab_host="Top10F"
 /clone_lib="S21SNU520"
 /note="Organ: Stomach; Vector: pTZ18RP1; Site_1: EcoRI;
 Site_2: NotI; The poly (A) RNA was dephosphorylated with

ORIGIN

Alignment Scores:
 Pred. No.: 0.00217 Length: 604
 Score: 113.00 Matches: 21
 Percent Similarity: 100.0% Conservative: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 100.0% Indels: 0
 DB: 3 Gaps: 0

US-09-606-910E-1 (1-21) x BM791359 (1-604)

Qy 1 AlahisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
 |||||
 Db 124 GCCCAGCGTGTACCTCGGCCCGACACCGCGCGCGCTCCACCGCCCCCA 183
 Qy 21 Ala 21
 ||||
 Db 184 GCC 186

RESULT 3
 DR422781
 LOCUS DR422781 754 bp mRNA linear EST 29-JUN-2005
 DEFINITION nav17d09.y1 Human pterygium. Unnormalized (nav) Homo sapiens cDNA
 clone nav17d09 5', mRNA sequence.

ACCESSION DR422781
 VERSION DR422781.1 GI:68324797
 KEYWORDS EST.

SOURCE

ORGANISM Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Hominidae; Homo.

REFERENCE 1 (bases 1 to 754)
 AUTHORS Wistow,G., Jaworski,C., Aryankalayil-John,M., Rowsay,J.J., Cox,C.,
 Reid,T., Dushku,N. and Carper, D.

TITLE NEIBank analysis of Human pterygium
 JOURNAL Unpublished (2005)
 COMMENT Contact: Wistow G
 Section on Molecular Structure and Function
 National Eye Institute

6/331, NIH, Bethesda, MD 20892-2740, USA
 Tel: 301 402 3452
 Fax: 301 496 0078
 Email: graeme@helix.nih.gov
 Plate: 17 row: d column: 09
 Seq primer: Universal M13 Reverse.

FEATURES

source

1..754
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="nav17d09"
 /tissue_type="Pterygium"
 /dev_stage="Adult"
 /lab_host="EMDH10B"
 /note="Organ: Eye; Vector: pCMVSPORT6; RNA was extracted
 from 9 pooled human pterygia. A directionally cloned cDNA

bacterial alkaline phosphatase (BAP) and then decapped with tabacco acid pyrophosphatase (TAP). The decapped intact mRNA was ligated with DNA-RNA linker including EcoR I site by treatment of 14 RNA ligase and the first strand cDNA was synthesized from oligo dT-selected mRNA by priming with dT-tailed vector. The dT-tailed vector was adjusted to have about 60nt. The cDNA vector was circularized with E. coli DNA ligase after digestion of EcoRI which site is also included in vector. An RNA strand converted to a DNA strand by Okayama-Berg method. The obtained cDNA vectors were used for transformation of competent cells E. coli Top10F' by electroporation method. The cDNA libraries constructed by this method are full-length enriched cDNA library."

library in the pCMVSPORT6 vector (Invitrogen) was constructed at Bioserve Biotechnology (Laurel MD) essentially following the protocols of the SuperScript Plasmid System, full details of which are contained in the manufacturer's instruction manual (http://www.lifetech.com/). First strand synthesis was carried out using a Not I primer-adaptor [5'-pGACTAGTTAGATCGGCGGCCG(T)15-3']. cDNA was cloned in Not I/Sal I sites. EST analysis was performed at the NIH Intramural Sequencing Center (NISC). Analyzed data available through http://neibank.nei.nih.gov."

ORIGIN

Alignment Scores: 0.00269 Length: 754
 Pred. No.: 113.00 Matches: 21
 Score: 113.00
 Percent Similarity: 100.0% Conservative: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 100.0% Indels: 0
 DB: 8 Gaps: 0

US-09-606-910E-1 (1-21) x DR422781 (1-754)

Qy 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
 |||||
 Db 54 GCCCAGCGTGTCACTCGGCCCGGACACCGCGCGGGCTCCACCGCACCCCA 113
 |||||
 Qy 21 Ala 21
 |||||
 Db 114 GCC 116

RESULT 4
 BUS42454 877 bp mRNA linear EST 13-SEP-2002
 LOCUS AGENCOURT_10322173 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6574322
 5', mRNA sequence.
 ACCESSION BUS42454
 VERSION BUS42454.1 GI:22852937
 KEYWORDS EST.
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Hominidae; Homo.

REFERENCE 1 (bases 1 to 877)
 AUTHORS NIH-MGC http://mgs.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov

Tissue Procurement: DCTD/DRP
 cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov

Plate: L1C82769 row: h column: 02
 High quality sequence stop: 760.
 Location/Qualifiers

FEATURES

source

1..877
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6574322"
 /tissue_type="carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH MGC 40"
 /note="organ: prostate; Vector: pOTB7; Site 1: XhoI;
 Site 2: EcoRI; cDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCACGAG(G). Library constructed by
 Ling Hong in the laboratory of Gerald M. Rubin (University

ORIGIN

Alignment Scores: 0.00311 Length: 877
 Pred. No.: 113.00 Matches: 21
 Score: 113.00
 Percent Similarity: 100.0% Conservative: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 100.0% Indels: 0
 DB: 5 Gaps: 0

US-09-606-910E-1 (1-21) x BU542454 (1-877)

Qy 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
 |||||
 Db 53 GCCCAGCGTGTCACTCGGCCCGGACACCGCGCGGGCTCCACCGCCCCCA 112
 |||||
 Qy 21 Ala 21
 |||||
 Db 113 GCC 115

RESULT 5
 CA489836 959 bp mRNA linear EST 14-NOV-2002
 LOCUS AGENCOURT_10810668 MAPCL Homo sapiens cDNA clone IMAGE:6722324 5',
 mRNA sequence.

ACCESSION CA489836
 VERSION CA489836.1 GI:24952627
 KEYWORDS EST.
 SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Hominidae; Homo

REFERENCE 1 (bases 1 to 959)
 AUTHORS NIH-MGC http://mgs.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov
 Tissue Procurement: Kristi A. Eglund, Ira Pastan
 cDNA Library Preparation: Invitrogen Corp
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov

Plate: L1A414284 row: n column: 20
 High quality sequence start: 31
 High quality sequence stop: 446.

FEATURES

source

1..959
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6722324"
 /cell_line="ZR-75-1, MCF7, SK-BR-3, MDA-MB-231,
 hTERT-HMEL LNCap"
 /lab_host="EMDH10B"
 /clone_lib="MAPCL"
 /note="vector: pCMV-SPORT6; Site 1: EcoRV; Site 2: Not I;
 Subtracted with brain, liver, lung, kidney and muscle.
 Directionally cloned. Priming method: oligo-dT. Average
 insert size: 1800 bp. Library amplification: 26,000 fold.
 Kristi A. Eglund, James J. Vincent, Robert Strausberg,
 Bungkok Lee & Ira Pastan. Discovery of new breast
 cancer genes encoding membrane and secreted proteins.
 Manuscript submitted."

ORIGIN

Alignment Scores:

```

Pred. No.: 0.00339 Length: 959
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 6 Gaps: 0

US-09-606-910E-1 (1-21) x CA489836 (1-959)

Qy 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db 427 GCCCAGCGTGTCACTCTGCCCGGACACACAGCGCGCGCGGCTCCACCGCCCCCA 486
Qy 21 Ala 21
Db 487 GCC 489

RESULT 6
BU148487 1113 bp mRNA linear EST 03-SEP-2002
DEFINITION AGENCOURT 8670479 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6380642
5', mRNA sequence.
ACCESSION BU148487
VERSION BU148487.1 GI:22662019
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE 1 (bases 1 to 1113)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DRP
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2569 row: j column: 03
High quality sequence stop: 235.
FEATURES
Location/Qualifiers
1..1113
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6380642"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_40"
/note="Organ: prostate; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
ORIGIN
Alignment Scores:
Pred. No.: 0.00392 Length: 1113
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 5 Gaps: 0

US-09-606-910E-1 (1-21) x BU148487 (1-1113)

Pred. No.: 0.00339 Length: 959
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 6 Gaps: 0

US-09-606-910E-1 (1-21) x CA489836 (1-959)

Qy 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db 427 GCCCAGCGTGTCACTCTGCCCGGACACACAGCGCGCGGCTCCACCGCCCCCA 486
Qy 21 Ala 21
Db 487 GCC 489

RESULT 7
BU542996 1130 bp mRNA linear EST 13-SEP-2002
LOCUS AGENCOURT 10338707 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6574979
DEFINITION 5', mRNA sequence.
ACCESSION BU542996
VERSION BU542996.1 GI:22853479
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE 1 (bases 1 to 1130)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DRP
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2771 row: c column: 11
High quality sequence start: 27
High quality sequence stop: 246.
FEATURES
Location/Qualifiers
1..1130
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6574979"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_40"
/note="Organ: prostate; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
ORIGIN
Alignment Scores:
Pred. No.: 0.00398 Length: 1130
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 5 Gaps: 0

US-09-606-910E-1 (1-21) x BU542996 (1-1130)

Qy 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db 87 GCCCAGCGTGTCACTCTGCCCGGACACACAGCGCGCGGCTCCACCGCCCCCA 146
Qy 21 Ala 21
Db 108 GCC 110

RESULT 7
BU542996 1130 bp mRNA linear EST 13-SEP-2002
LOCUS AGENCOURT 10338707 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6574979
DEFINITION 5', mRNA sequence.
ACCESSION BU542996
VERSION BU542996.1 GI:22853479
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE 1 (bases 1 to 1130)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DRP
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2771 row: c column: 11
High quality sequence start: 27
High quality sequence stop: 246.
FEATURES
Location/Qualifiers
1..1130
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6574979"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_40"
/note="Organ: prostate; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
ORIGIN
Alignment Scores:
Pred. No.: 0.00398 Length: 1130
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 5 Gaps: 0

US-09-606-910E-1 (1-21) x BU542996 (1-1130)

Qy 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db 87 GCCCAGCGTGTCACTCTGCCCGGACACACAGCGCGCGGCTCCACCGCCCCCA 146
Qy 21 Ala 21
Db 108 GCC 110

```

```

Db          147 GCC 149
RESULT 8
BQ936898
LOCUS
DEFINITION AGENCOURT_8919110 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6383855
5', mRNA sequence.
ACCESSION BQ936898
VERSION BQ936898.1 GI:22352281
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 1234)
NIH-MGC http://mgs.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DTF
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2577 row: 0 column: 24
High quality sequence stop: 245.
Location/Qualifiers
1..1234
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6383855"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_40"
/notes="Organ: prostate; Vector: pOTB7; Site 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
ORIGIN
Alignment Scores:
Pred. No.: 0.00433 Length: 1234
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
Gaps: 0
DB: 5

US-09-606-910E-1 (1-21) x BQ936898 (1-1234)

QY 1 AlaHisGlyValThrSerAlaProArgProAlaProGlySerThrAlaProPro 20
|||||
Db 48 GCCACGGTGTACCTCTGCCCGGACACAGCGCGCGGCTCCACCGCCCCCA 107
|||||
QY 21 Ala 21
|||
Db 108 GCC 110

RESULT 9
BQ935496
LOCUS
DEFINITION AGENCOURT_8776475 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6481609
5', mRNA sequence.
ACCESSION BQ935496
VERSION BQ935496.1 GI:22334753
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 1262)
NIH-MGC http://mgs.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DTF
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2573 row: f column: 09
High quality sequence start: 46
High quality sequence stop: 157.
Location/Qualifiers
1..1262
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6382088"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_40"
/notes="Organ: prostate; Vector: pOTB7; Site 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
ORIGIN
Alignment Scores:
Pred. No.: 0.00443 Length: 1262
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
Gaps: 0
DB: 5

US-09-606-910E-1 (1-21) x BQ935496 (1-1262)

QY 1 AlaHisGlyValThrSerAlaProArgProAlaProGlySerThrAlaProPro 20
|||||
Db 74 GCCACGGTGTACCTCTGCCCGGACACAGCGCGGCTCCACCGCCCCCA 133
|||||
QY 21 Ala 21
|||
Db 134 GCC 136

RESULT 10
BQ920055
LOCUS
DEFINITION AGENCOURT_10031674 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6481609
5', mRNA sequence.
ACCESSION BQ920055
VERSION BQ920055.1 GI:22334753
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.

```

```

REFERENCE
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/.
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL       Unpublished (1999)
COMMENT       Contact: Robert Strausberg, Ph.D.
              Email: c9apbs-r@mail.nih.gov
              Tissue Procurement: DCTD/DTP
              cDNA Library Preparation: Rubin Laboratory
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA sequencing by: Agencourt Bioscience Corporation
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LICM2663 row: e column: 02
              High quality sequence start: 56
              High quality sequence stop: 237.
              Location/Qualifiers
                1. 1343
                  /organism="Homo sapiens"
                  /mol_type="mRNA"
                  /db_xref="taxon:9606"
                  /clone="IMAGE:6481609"
                  /tissue_type="carcinoma, cell line"
                  /lab_host="DH10B (phage-resistant)"
                  /clone_lib="NIH MGC 40"
                  /note="Organ: prostate; Vector: pOTB7; Site 1: XhoI;
                  Site 2: EcoRI; cDNA made by oligo-dT priming.
                  Directionally cloned into EcoRI/XhoI sites using the
                  following 5' adaptor: GGCACGAG(G). Library constructed by
                  Ling Hong in the laboratory of Gerald M. Rubin (University
                  of California, Berkeley) using ZAP-cDNA synthesis kit
                  (Stratagene) and Superscript II RT (Life Technologies).
                  Note: this is a NIH MGC Library."
FEATURES
source

```

```

ORIGIN
Alignment Scores:
Pred. No.: 0.0047 Length: 1343
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 5 Gaps: 0

US-09-606-910E-1 (1-21) x BQ20055 (1-1343)

Qy 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db 90 GCCCACGGTGTACCTCTGCCCGGACACAGGCGCGCGGCTCCACGCGCCCCCA 149
Qy 21 Ala 21
Db 150 GCC 152

RESULT 11
BUI52566 1349 bp mRNA linear EST 03-SEP-2002
LOCUS AGENCOURT 8670683 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6380559
DEFINITION 5', mRNA sequence.
ACCESSION BUI52566
VERSION BUI52566.1 GI:22666098
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
1 (bases 1 to 1349)
NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTF

```

```

cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2569 row: f column: 16
High quality sequence stop: 291.
Location/Qualifiers
1. .1349
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6380559"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 40"
/note="Organ: prostate; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:
Pred. No.: 0.00472 Length: 1349
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservativeness: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
na. 5 Gaps: 0

```

```

US-09-606-910E-1 (1-21) x BUI52566 (1-1349)

Qy 1 AlahisglyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
    |||||
Db 37 GCCACCGGTGTACCTCGGCCCGGACACAGCGCGGCCCGGGCTCCACGCGCCCCCA 96
    |||||

Qy 21 Ala 21
    |||
Db 97 GCC 99

RESULT 12
BU542790 1420 bp mRNA linear EST 13-SEP-2002
LOCUS AGENCOURT_10334841 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6574725
DEFINITION 5', mRNA sequence.
ACCESSION BU542790
VERSION BU542790.1 GI:22853273
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE 1 (bases 1 to 1420)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/BTP
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2770 row: h column: 21
High quality sequence stop: 288.

```

FEATURES
source

Location/Qualifiers
1. .1420
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6574725"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_40"
/note="Organ: prostate; Vector: pOTB7; Site:1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dt priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:
Pred. No.: 0.00496 Length: 1420
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 5 Gaps: 0

US-09-606-910E-1 (1-21) x BU542790 (1-1420)

QY 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
|||||
Db 37 GCCCAGCGGTGTCACCTCGCGCCGGACACAGCGCGCGGCTCCACCGCCCCCA 96
QY 21 Ala 21
|||
Db 97 GCC 99

RESULT 13

BU543309 1531 bp mRNA linear EST 13-SRP-2002
LOCUS AGENCOURT_10327072 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6575349
DEFINITION 5', mRNA sequence.

ACCESSION BU543309
VERSION BU543309.1 GI:22853792
KEYWORDS EST.
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.

REFERENCE 1 (bases 1 to 1531)
NIH-MGC http://mgs.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP

cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov
Plate: LCM2772 row: b column: 21
High quality sequence stop: 166.

FEATURES
source

Location/Qualifiers
1. .1531
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6575349"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"

/clone_lib="NIH_MGC_40"
/note="Organ: prostate; Vector: pOTB7; Site:1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dt priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:
Pred. No.: 0.00533 Length: 1531
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 5 Gaps: 0

US-09-606-910E-1 (1-21) x BU543309 (1-1531)

QY 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
|||||
Db 48 GCCCAGCGGTGTCACCTCGCGCCGGACACAGCGCGGCTCCACCGCCCCCA 107

QY 21 Ala 21
|||
Db 108 GCC 110

RESULT 14

BU543354 1268 bp mRNA linear EST 21-AUG-2002
LOCUS AGENCOURT_8777815 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6384308
DEFINITION 5', mRNA sequence.

ACCESSION BU543354
VERSION BU543354.1 GI:22359032
KEYWORDS EST.
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.

REFERENCE 1 (bases 1 to 1268)
NIH-MGC http://mgs.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP

cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov
Plate: LCM2579 row: b column: 21
High quality sequence stop: 177.

FEATURES
source

Location/Qualifiers
1. .1268
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6384308"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_40"

/note="Organ: prostate; Vector: pOTB7; Site:1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dt priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: April 14, 2006, 18:54:01 ; Search time 122.415 Seconds
(without alignments)
304.937 Million cell updates/sec

Title: US-09-606-910E-1
Perfect score: 113
Sequence: 1 AHGVTGAPTRDPAGSTAPPA 21

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q=/abes/ABSWEB.spool/US09606910/runat_14042006_091623_1610/app_query.fasta_1
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-HOST=abs808 -USER=US09606910@cgn_1_1_237 @runat_14042006_091623_1610 -NCPU=6
-ICPU=3 -NO MMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/1/ina/1 COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5 COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/1/ina/H COMB.seq.*
6: /cgn2_6/ptodata/1/ina/PTUS COMB.seq.*
7: /cgn2_6/ptodata/1/ina/PE COMB.seq.*
8: /cgn2_6/ptodata/1/ina/RE COMB.seq.*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	113	100.0	1721	3	US-10-029-517-3
2	113	100.0	1804	3	Sequence 3, Appli
3	113	100.0	8186	3	Sequence 17, Appl
4	105	92.9	572	3	US-10-029-517-19
5	101	89.4	519	3	Sequence 18, Appl
6	101	89.4	534	3	US-10-029-517-18
7	98	86.7	518	3	Sequence 42, Appl
8	98	86.7	3343	3	US-09-646-028-42
9	95	84.1	6192	2	Sequence 46, Appl
					Sequence 101, App
					Sequence 102, App
					Sequence 1, Appli

10	95	84.1	6192	3	US-09-083-116-1	Sequence 1, Appli
11	95	84.1	6192	3	US-09-134-916A-1	Sequence 1, Appli
12	95	84.1	6449	2	US-08-479-537A-4	Sequence 4, Appli
13	95	84.1	6449	3	US-09-083-116-4	Sequence 4, Appli
14	95	84.1	6449	3	US-09-134-916A-4	Sequence 4, Appli
15	94	83.2	981	3	US-10-029-517-16	Sequence 16, Appl
16	85	75.2	68	3	US-09-304-967-51	Sequence 51, Appl
17	84	74.3	83	3	US-09-304-967-75	Sequence 75, Appl
18	83	73.5	48	3	US-09-304-967-47	Sequence 47, Appl
19	83	73.5	48	3	US-09-304-967-67	Sequence 67, Appl
20	83	73.5	48	3	US-09-304-967-90	Sequence 90, Appl
21	83	73.5	68	3	US-09-304-967-49	Sequence 49, Appl
22	83	73.5	68	3	US-09-304-967-53	Sequence 53, Appl
23	83	73.5	68	3	US-09-304-967-55	Sequence 55, Appl
24	83	73.5	68	3	US-09-304-967-57	Sequence 57, Appl
25	83	73.5	78	3	US-09-304-967-92	Sequence 92, Appl
26	83	73.5	78	3	US-09-304-967-94	Sequence 94, Appl
27	83	73.5	78	3	US-09-304-967-96	Sequence 96, Appl
28	83	73.5	78	3	US-09-304-967-98	Sequence 98, Appl
29	83	73.5	78	3	US-09-304-967-100	Sequence 100, App
30	83	73.5	78	3	US-09-304-967-102	Sequence 102, App
31	83	73.5	83	3	US-09-304-967-69	Sequence 69, Appl
32	83	73.5	83	3	US-09-304-967-71	Sequence 71, Appl
33	83	73.5	83	3	US-09-304-967-73	Sequence 73, Appl
34	83	73.5	83	3	US-09-304-967-77	Sequence 77, Appl
35	83	73.5	83	3	US-09-304-967-79	Sequence 79, Appl
36	65	57.5	60	3	US-09-475-947A-246	Sequence 246, App
37	64.5	57.1	11220	3	US-09-105-537-32	Sequence 32, Appl
38	64.5	57.1	36778	3	US-09-105-537-5	Sequence 5, Appli
39	64.5	57.1	38506	3	US-09-320-878-19	Sequence 19, Appl
40	64.5	57.1	38506	3	US-09-141-908-1	Sequence 1, Appli
41	64.5	57.1	38506	3	US-09-657-440-19	Sequence 19, Appl
42	64.5	57.1	38506	3	US-09-793-708-19	Sequence 19, Appl
43	63.5	56.2	4403765	3	US-09-103-840A-2	Sequence 2, Appli
44	63.5	56.2	4411529	3	US-09-103-840A-1	Sequence 1, Appli
45	61	54.0	942	3	US-09-902-540-7357	Sequence 7357, Ap

ALIGNMENTS

RESULT 1

US-10-029-517-3
; Sequence 3, Application US/10029517
; Patent No. 6716627
; GENERAL INFORMATION:
; APPLICANT: Kenneth W. Dobie
; APPLICANT: Susan J. Myers
; TITLE OF INVENTION: ANTISENSE MODULATION OF MUCIN 1, TRANSMEMBRANE EXPRESSION
; FILE REFERENCE: RTS-0352
; CURRENT APPLICATION NUMBER: US/10/029,517
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 107
; SEQ ID NO 3
; LENGTH: 1721
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (58)...(1605)
US-10-029-517-3

Alignment Scores:
Pred. No.: 0.000239 Length: 1721
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 3 Gaps: 0

US-09-606-910E-1 (1-21) x US-10-029-517-3 (1-1721)

Qy 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
|||||

```
Db 580 GCCACGGTGTACCTCGGCCCGGACACACGAGCGCGCCCGGGCTCCACCGCCCCCCCCA 639
Qy 21 Ala 21
Db 640 GCC 642

RESULT 2
US-10-029-517-17
; Sequence 17, Application US/10029517
; Patent No. 6716627
; GENERAL INFORMATION:
; APPLICANT: Kenneth W. Dobie
; APPLICANT: Susan J. Myers
; TITLE OF INVENTION: ANTISENSE MODULATION OF MUCIN 1, TRANSMEMBRANE EXPRESSION
; FILE REFERENCE: RTS-0352
; CURRENT APPLICATION NUMBER: US/10/029,517
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 107
; SEQ ID NO 17
; LENGTH: 1804
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (73)...(1500)
US-10-029-517-17

Alignment Scores:
Pred. No.: 0.000251 Length: 1804
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 3 Gaps: 0

US-09-606-910E-1 (1-21) x US-10-029-517-17 (1-1804)
Qy 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db 475 GCCACGGTGTACCTCGGCCCGGACACACGAGCGCGCCCGGGCTCCACCGCCCCCCCCA 534
Qy 21 Ala 21
Db 535 GCC 537

RESULT 3
US-10-029-517-19
; Sequence 19, Application US/10029517
; Patent No. 6716627
; GENERAL INFORMATION:
; APPLICANT: Kenneth W. Dobie
; APPLICANT: Susan J. Myers
; TITLE OF INVENTION: ANTISENSE MODULATION OF MUCIN 1, TRANSMEMBRANE EXPRESSION
; FILE REFERENCE: RTS-0352
; CURRENT APPLICATION NUMBER: US/10/029,517
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 107
; SEQ ID NO 19
; LENGTH: 8186
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 6899
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 7155
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 7184
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 7957
```

```
; OTHER INFORMATION: unknown
; NAME/KEY: intron
; LOCATION: (2997)...(3498)
; OTHER INFORMATION: intron 1
; NAME/KEY: intron:exon junction
; LOCATION: (3498)...(3499)
; OTHER INFORMATION: intron 1:exon 2
; NAME/KEY: exon
; LOCATION: (3508)...(3599)
; OTHER INFORMATION: exon 2d
; NAME/KEY: exon:intron junction
; LOCATION: (3982)...(3983)
; OTHER INFORMATION: exon 2a:intron 2a
; NAME/KEY: intron:exon junction
; LOCATION: (4205)...(4206)
; OTHER INFORMATION: intron 2c:exon 3c
; NAME/KEY: intron:exon junction
; LOCATION: (4259)...(4260)
; OTHER INFORMATION: intron 2d:exon 3d
; NAME/KEY: exon
; LOCATION: (4260)...(4328)
; OTHER INFORMATION: exon 3d
; NAME/KEY: intron:exon junction
; LOCATION: (4632)...(4633)
; OTHER INFORMATION: intron 3:exon 4
; NAME/KEY: exon
; LOCATION: (4914)...(5035)
; OTHER INFORMATION: exon 5
; NAME/KEY: intron
; LOCATION: (5266)...(6293)
; OTHER INFORMATION: intron 6
US-10-029-517-19

Alignment Scores:
Pred. No.: 0.00116 Length: 8186
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 3 Gaps: 0

US-09-606-910E-1 (1-21) x US-10-029-517-19 (1-8186)
Qy 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db 3843 GCCACGGTGTACCTCGGCCCGGACACACGAGCGCGCCCGGGCTCCACCGCCCCCCCCA 3902
Qy 21 Ala 21
Db 3903 GCC 3905

RESULT 4
US-10-029-517-18
; Sequence 18, Application US/10029517
; Patent No. 6716627
; GENERAL INFORMATION:
; APPLICANT: Kenneth W. Dobie
; APPLICANT: Susan J. Myers
; TITLE OF INVENTION: ANTISENSE MODULATION OF MUCIN 1, TRANSMEMBRANE EXPRESSION
; FILE REFERENCE: RTS-0352
; CURRENT APPLICATION NUMBER: US/10/029,517
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 107
; SEQ ID NO 18
; LENGTH: 572
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (67)...(572)
US-10-029-517-18

Alignment Scores:
```


Pred. No.: 0.000653 Length: 572
Score: 105.00 Matches: 20
Percent Similarity: 95.2% Conservative: 0
Best Local Similarity: 95.2% Mismatches: 1
Query Match: 92.9% Indels: 0
DB: 3 Gaps: 0

US-09-606-910E-1 (1-21) x US-10-029-517-18 (1-572)

QY 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
DB 496 GCCACGGGTCTACCTCGCGCCCGGACACAGCCGCGCGCTCCACCGCCCGCCCA 555

QY 21 Ala 21

DB 556 GCC 558

RESULT 5

US-09-646-028-42
; Sequence 42, Application US/09646028
; Patent No. 6562347
; GENERAL INFORMATION:
; APPLICANT: Kwak, Larry
; APPLICANT: Biragyn, Arya
; TITLE OF INVENTION: METHODS AND COMPOSITIONS OF
; TITLE OF INVENTION: CHEMOKINE-TUMOR ANTIGEN FUSION PROTEINS AS CANCER VACCINES
; FILE REFERENCE: 14014.0316/P
; CURRENT APPLICATION NUMBER: US/09/646,028
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 60/077,745
; PRIOR FILING DATE: 1998-03-12
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 42
; LENGTH: 519
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of artificial sequence:/note=synthetic construct
US-09-646-028-42

Alignment Scores:
Pred. No.: 0.00171 Length: 519
Score: 101.00 Matches: 19
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 89.4% Indels: 0
DB: 3 Gaps: 0

US-09-606-910E-1 (1-21) x US-09-646-028-42 (1-519)

QY 3 GlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 21
DB 241 GGTGTAACCTCTGCGCCCGGACACTCGCCCGGCGGTCTACTGCTCCGCGGCA 297

RESULT 6

US-09-646-028-46
; Sequence 46, Application US/09646028
; Patent No. 6562347
; GENERAL INFORMATION:
; APPLICANT: Kwak, Larry
; APPLICANT: Biragyn, Arya
; TITLE OF INVENTION: METHODS AND COMPOSITIONS OF
; TITLE OF INVENTION: CHEMOKINE-TUMOR ANTIGEN FUSION PROTEINS AS CANCER VACCINES
; FILE REFERENCE: 14014.0316/P
; CURRENT APPLICATION NUMBER: US/09/646,028
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 60/077,745
; PRIOR FILING DATE: 1998-03-12
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 46
; LENGTH: 514

; TYPE: DNA
; ORGANISM: Artificial Sequence

; FEATURE:
; OTHER INFORMATION: Description of artificial sequence:/note=synthetic construct
US-09-646-028-46

Alignment Scores:

Pred. No.: 0.00176 Length: 534
Score: 101.00 Matches: 19
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 89.4% Indels: 0
DB: 3 Gaps: 0

US-09-606-910E-1 (1-21) x US-09-646-028-46 (1-534)

QY 3 GlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 21

DB 256 GGTGTAACCTCTGCGCCCGGACACTCGCCCGGCGGTCTACTGCTCCGCGGCA 312

RESULT 7

US-10-029-517-101
; Sequence 101, Application US/10029517
; Patent No. 6716627
; GENERAL INFORMATION:
; APPLICANT: Kenneth W. Dobie
; APPLICANT: Susan J. Myers
; TITLE OF INVENTION: ANTISENSE MODULATION OF MUCIN 1, TRANSMEMBRANE EXPRESSION
; FILE REFERENCE: RTS-0352
; CURRENT APPLICATION NUMBER: US/10/029,517
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 107
; SEQ ID NO 101
; LENGTH: 518
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Description of artificial sequence:/note=synthetic construct
US-10-029-517-101

Alignment Scores:
Pred. No.: 0.00378 Length: 518
Score: 98.00 Matches: 18
Percent Similarity: 90.5% Conservative: 1
Best Local Similarity: 85.7% Mismatches: 2
Query Match: 86.7% Indels: 0
DB: 3 Gaps: 0

US-09-606-910E-1 (1-21) x US-10-029-517-101 (1-518)

QY 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
DB 420 GCCACGATGTCACTCGCCCGGACACACAGCCGCGGTCTCCACCGCCCGCCA 479

QY 21 Ala 21

DB 480 GCC 482

RESULT 8

US-10-029-517-102
; Sequence 102, Application US/10029517
; Patent No. 6716627
; GENERAL INFORMATION:
; APPLICANT: Kenneth W. Dobie
; APPLICANT: Susan J. Myers
; TITLE OF INVENTION: ANTISENSE MODULATION OF MUCIN 1, TRANSMEMBRANE EXPRESSION
; FILE REFERENCE: RTS-0352
; CURRENT APPLICATION NUMBER: US/10/029,517
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 107
; SEQ ID NO 102
; LENGTH: 3343
; TYPE: DNA
; ORGANISM: Homo sapiens

```

; FEATURE:
US-10-029-517-102

Alignment Scores:
Pred. No.: 0.0251 Length: 3343
Score: 98.00 Matches: 18
Percent Similarity: 90.5% Conservative: 1
Best Local Similarity: 85.7% Mismatches: 2
Query Match: 86.7% Indels: 0
DB: 3 Gaps: 0

US-09-606-910E-1 (1-21) x US-10-029-517-102 (1-3343)

QY 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db 1626 GCCCAGCATGTCACTCAGCCCGGACAAACAGCCAGCCCGGGCTCCACCGCCCCCA 1685

QY 21 Ala 21
Db 1686 GCC 1688

RESULT 9
US-08-479-537A-1
; Sequence 1, Application US/08479537A
; Patent No. 5861381
; GENERAL INFORMATION:
; APPLICANT: CHAMBON, Pierre
; APPLICANT: KIENY, Marie-Paule
; APPLICANT: LATHE, Richard
; APPLICANT: HAREUVENI, Mara
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
; TITLE OF INVENTION: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/479,537A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 90/13101
; FILING DATE: 23-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/FR91/00835
; FILING DATE: 23-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/039,320
; FILING DATE: 04-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/403,576
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Teskin, Robin L.
; REGISTRATION NUMBER: 35,030
; REFERENCE/DOCKET NUMBER: 017753-025
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6192 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; FEATURE:
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 58..120
; FEATURE:
; NAME/KEY: repeat_region
; LOCATION: 439..5239
; OTHER INFORMATION: /note= "The nucleotides spanning
; OTHER INFORMATION: 439-5239 constitute a repeated region wherein the repeat is 6
; OTHER INFORMATION: nucleotides and encodes 20 amino acids, 17 of which are fixed
; OTHER INFORMATION: The number of such repeats varies from 1 to 80."
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 121..6166
; FEATURE:
; NAME/KEY: repeat_region
; LOCATION: 457
; OTHER INFORMATION: /note= "Nucleotide 457 is X1 = NNN
; OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
; OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
; FEATURE:
; NAME/KEY: repeat_region
; LOCATION: 487
; OTHER INFORMATION: /note= "Nucleotide 487 is Y = NNN
; OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA
; OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
; FEATURE:
; NAME/KEY: repeat_region
; LOCATION: 496
; OTHER INFORMATION: /note= "Nucleotide 496 is X2 = NNN
; OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
; OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
US-08-479-537A-1

Alignment Scores:
Pred. No.: 0.104 Length: 6192
Score: 95.00 Matches: 18
Percent Similarity: 85.7% Conservative: 0
Best Local Similarity: 85.7% Mismatches: 3
Query Match: 84.1% Indels: 0
DB: 2 Gaps: 0

US-09-606-910E-1 (1-21) x US-08-479-537A-1 (1-6192)

QY 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db 460 GCCCAGGTGTCACTCGCCCCCGACNNNAGCCGCTCCACCGCCCCNNN 519

QY 21 Ala 21
Db 520 GCC 522

RESULT 10
US-09-083-116-1
; Sequence 1, Application US/09083116
; Patent No. 6203795
; GENERAL INFORMATION:
; APPLICANT: CHAMBON, Pierre
; APPLICANT: KIENY, Marie-Paule
; APPLICANT: LATHE, Richard
; APPLICANT: HAREUVENI, Mara
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
; TITLE OF INVENTION: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/083,116
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/479,537
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-6620
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 6192 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 58..120
FEATURE:
NAME/KEY: repeat_region
LOCATION: 439..5239
OTHER INFORMATION: /note= "The nucleotides spanning 439-5239 constitute a repeated region wherein the repeat is 6 nucleotides and encodes 20 amino acids, 17 of which are fixed
OTHER INFORMATION: The number of such repeats varies from 1 to 80."
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 121..6166
FEATURE:
NAME/KEY: repeat_region
LOCATION: 457
OTHER INFORMATION: /note= "Nucleotide 457 is X1 = NNN
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
OTHER INFORMATION: or CCG, and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: repeat_region
LOCATION: 487
OTHER INFORMATION: /note= "Nucleotide 487 is Y = NNN
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA
OTHER INFORMATION: or ACG, and Asn = AAT or AAC."
FEATURE:
NAME/KEY: repeat_region
LOCATION: 496
OTHER INFORMATION: /note= "Nucleotide 496 is X2 = NNN
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
OTHER INFORMATION: or CCG, and Ala = GCT, GCC, GCA, or GCG."
US-09-083-116-1
Alignment Scores:
Pred. No.: 0.104 Length: 6192
Score: 95.00 Matches: 18
Percent Similarity: 85.7% Conservative: 0
Best Local Similarity: 85.7% Mismatches: 3
Query Match: 84.1% Indels: 0

DB: 3 Gaps: 0
US-09-606-910E-1 (1-21) x US-09-083-116-1 (1-6192)
Qy 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db 460 GCCACCGGTGTCACTCGGCCCGGACNNAGCCGNNCCGGGCTCCACCGCCCCNNN 519
Qy 21 Ala 21
Db 520 GCC 522
RESULT 11
US-09-134-916A-1
; Sequence 1, Application US/09134916A
; Patent No. 6328956
; GENERAL INFORMATION:
; APPLICANT: CHAMBON, Pierre
; APPLICANT: KIENY, Marie-Paule
; APPLICANT: LATHE, Richard
; APPLICANT: HAREUVENI, Mara
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
; TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESS: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/134,916A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/479,537
; FILING DATE: 07-JUN-1995
; APPLICATION NUMBER: FR 90/13101
; FILING DATE: 23-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/FR91/00835
; FILING DATE: 23-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/039,320
; FILING DATE: 04-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/403,576
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Teskin, Robin L.
; REGISTRATION NUMBER: 35,030
; REFERENCE/DOCKET NUMBER: 017753-025
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6192 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 58..120
; FEATURE:
; NAME/KEY: repeat_region

```
;
; LOCATION: 439..5239
; OTHER INFORMATION: /note= "The nucleotides spanning
; OTHER INFORMATION: 439-5239 constitute a repeated region wherein the repeat is 6
; OTHER INFORMATION: nucleotides and encodes 20 amino acids, 17 of which are fixed
; OTHER INFORMATION: The number of such repeats varies from 1 to 80."
;
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 121..6166
;
; NAME/KEY: repeat_region
; LOCATION: 457
; OTHER INFORMATION: /note= "Nucleotide 457 is X1 = NNN
; OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
; OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
;
; FEATURE:
; NAME/KEY: repeat_region
; LOCATION: 487
; OTHER INFORMATION: /note= "Nucleotide 487 is Y = NNN
; OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA
; OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
;
; FEATURE:
; NAME/KEY: repeat_region
; LOCATION: 496
; OTHER INFORMATION: /note= "Nucleotide 496 is X2 = NNN
; OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
; OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
;
; US-09-134-916A-1
;
; Alignment Scores:
; Pred. No.: 0.104 Length: 6192
; Score: 95.00 Matches: 18
; Percent Similarity: 85.7% Conservative: 0
; Best Local Similarity: 85.7% Mismatches: 3
; Query Match: 84.1% Indels: 0
; DB: 3 Gaps: 0
;
; US-09-606-910E-1 (1-21) x US-09-134-916A-1 (1-6192)
;
; QY 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
;
; Db 460 GCCCAGCGTGTCACTCGCGCCGACNNAGCCGNNCCGGCTCCACCGCCCCNN 519
;
; QY 21 Ala 21
;
; Db 520 GCC 522
;
; RESULT 12
; US-08-479-537A-4
; Sequence 4, Application US/08479537A
; Patent No. 5861381
;
; GENERAL INFORMATION:
; APPLICANT: CHAMBER, Pierre
; APPLICANT: KIENV, Marie-Paule
; APPLICANT: LATHE, Richard
; APPLICANT: HAREUVENI, Mara
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
; TITLE OF INVENTION: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/479,537A
; FILING DATE: 07-JUN-1995
```

```
;
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 90/13101
; FILING DATE: 23-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/FR91/00835
; FILING DATE: 23-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/039,320
; FILING DATE: 04-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/403,576
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Teskin, Robin L.
; REGISTRATION NUMBER: 35,030
; REFERENCE/DOCKET NUMBER: 017753-025
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6449 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 58..120
; FEATURE:
; NAME/KEY: repeat_region
; LOCATION: 439..5239
; OTHER INFORMATION: /note= "The nucleotides spanning
; OTHER INFORMATION: 439-5239 constitute a repeated region wherein the repeat is 6
; OTHER INFORMATION: nucleotides and encodes 20 amino acids, 17 of which are fixed
; OTHER INFORMATION: The number of such repeats varies from 1 to 80."
;
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 121..5661
; FEATURE:
; NAME/KEY: repeat_region
; LOCATION: 457
; OTHER INFORMATION: /note= "Nucleotide 457 is X1 = NNN
; OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
; OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
;
; FEATURE:
; NAME/KEY: repeat_region
; LOCATION: 487
; OTHER INFORMATION: /note= "Nucleotide 487 is Y = NNN
; OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA
; OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
;
; FEATURE:
; NAME/KEY: repeat_region
; LOCATION: 496
; OTHER INFORMATION: /note= "Nucleotide 496 is X2 = NNN
; OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
; OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
;
; US-08-479-537A-4
;
; Alignment Scores:
; Pred. No.: 0.108 Length: 6449
; Score: 95.00 Matches: 18
; Percent Similarity: 85.7% Conservative: 0
; Best Local Similarity: 85.7% Mismatches: 3
; Query Match: 84.1% Indels: 0
; DB: 2 Gaps: 0
;
; US-09-606-910E-1 (1-21) x US-08-479-537A-4 (1-6449)
;
; QY 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
;
; Db 460 GCCCAGCGTGTCACTCGCGCCGACNNAGCCGNNCCGGCTCCACCGCCCCNN 519
```

```
QY      21 Ala 21
Db      520 GCC 522

RESULT 13
US-09-083-116-4
; Sequence 4, Application US/09083116
; Patent No. 6203795
; GENERAL INFORMATION:
; APPLICANT: CHAMBERN, Pierre
; APPLICANT: KIENY, Marie-Paule
; APPLICANT: LATHE, Richard
; APPLICANT: HAREUVENI, Mara
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
; TITLE OF INVENTION: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/083,116
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/479,537
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/FR91/00835
; FILING DATE: 23-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/039,320
; FILING DATE: 04-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/403,576
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Teskin, Robin L.
; REGISTRATION NUMBER: 35,030
; REFERENCE/DOCKET NUMBER: 017753-025
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6449 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 58..120
; FEATURE:
; NAME/KEY: repeat_region
; LOCATION: 439..5239
; OTHER INFORMATION: /note= "The nucleotides spanning
; OTHER INFORMATION: 439-5239 constitute a repeated region wherein the repeat is 6
; OTHER INFORMATION: nucleotides and encodes 20 amino acids, 17 of which are fixed
; OTHER INFORMATION: The number of such repeats varies from 1 to 80."
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 121..5661
; FEATURE:

; NAME/KEY: repeat_region
; LOCATION: 457
; OTHER INFORMATION: /note= "Nucleotide 457 is X1 = NNN
; OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
; OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
; FEATURE:
; NAME/KEY: repeat_region
; LOCATION: 487
; OTHER INFORMATION: /note= "Nucleotide 487 is Y = NNN
; OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, AC
; OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
; FEATURE:
; NAME/KEY: repeat_region
; LOCATION: 496
; OTHER INFORMATION: /note= "Nucleotide 496 is X2 = NNN
; OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
; OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
US-09-083-116-4

Alignment Scores:
Pred. No.: 0.108 Length: 6449
Score: 95.00 Matches: 18
Percent Similarity: 85.7% Conservative: 0
Best Local Similarity: 85.7% Mismatches: 3
Query Match: 84.1% Indels: 0
DB: 3 Gaps: 0

US-09-606-910E-1 (1-21) x US-09-083-116-4 (1-6449)

QY      1 AlaHisGlyValThrSerAlaProGlyThrArgProAlaProGlySerThrAlaProPro 20
Db      460 GCCACGGTGTACCTCGGCCCGCCGACNNNAGCCGNNNCCGGGCTCCACGCCCCNNN 519

QY      21 Ala 21
Db      520 GCC 522

RESULT 14
US-09-134-916A-4
; Sequence 4, Application US/09134916A
; Patent No. 6328956
; GENERAL INFORMATION:
; APPLICANT: CHAMBERN, Pierre
; APPLICANT: KIENY, Marie-Paule
; APPLICANT: LATHE, Richard
; APPLICANT: HAREUVENI, Mara
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
; TITLE OF INVENTION: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/134,916A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/479,537
; FILING DATE: 07-JUN-1995
; APPLICATION NUMBER: FR 90/13101
; FILING DATE: 23-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/FR91/00835
; FILING DATE: 23-OCT-1991
```

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/039,320
; FILING DATE: 04-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/403,576
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Teskin, Robin L.
; REGISTRATION NUMBER: 35,030
; REFERENCE/DOCKET NUMBER: 017753-025
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6449 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 58..120
; FEATURE:
; NAME/KEY: repeat_region
; LOCATION: 439..5239
; OTHER INFORMATION: /note= "The nucleotides spanning
; 439-5239 constitute a repeated region wherein the repeat is 6
; nucleotides and encodes 20 amino acids, 17 of which are fixed
; and the number of such repeats varies from 1 to 80."
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 121..5661
; FEATURE:
; NAME/KEY: repeat_region
; LOCATION: 457
; OTHER INFORMATION: /note= "Nucleotide 457 is X1 = NNN
; which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
; or CCG; and Ala = GCT, GCC, GCA, or GCG."
; FEATURE:
; NAME/KEY: repeat_region
; LOCATION: 487
; OTHER INFORMATION: /note= "Nucleotide 487 is Y = NNN
; which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA
; or ACG; and Asn = AAT or AAC."
; FEATURE:
; NAME/KEY: repeat_region
; LOCATION: 496
; OTHER INFORMATION: /note= "Nucleotide 496 is X2 = NNN
; which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
; or CCG; and Ala = GCT, GCC, GCA, or GCG."
;
; US-09-134-916A-4
```

```
Alignment Scores:
Pred. No.: 0.108 Length: 6449
Score: 95.00 Matches: 18
Percent Similarity: 85.7% Conservative: 0
Best Local Similarity: 85.7% Mismatches: 3
Query Match: 84.1% Indels: 0
DB: 3 Gaps: 0
```

US-09-606-910E-1 (1-21) x US-09-134-916A-4 (1-6449)

```
QY 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db 460 GCCCAGGGTGTACCTGGCCCGGACNNNAGCCGNNNCCGGGCTCCACCGCCCCNNN 519
QY 21 Ala 21
Db 520 GCC 522
```

RESULT 15
US-10-029-517-16

```
; Sequence 16, Application US/10029517
; Patent No. 6716627
; GENERAL INFORMATION:
; APPLICANT: Kenneth W. Dobie
; APPLICANT: Susan J. Myers
; TITLE OF INVENTION: ANTISENSE MODULATION OF MUCIN 1, TRANSMEMBRANE EXPRESSION
; FILE REFERENCE: RPS-0352
; CURRENT APPLICATION NUMBER: US/10/029,517
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 107
; SEQ ID NO 16
; LENGTH: 981
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: exon:exon junction
; LOCATION: (464)...(465)
; OTHER INFORMATION: exon 3b:exon 4
;
; US-10-029-517-16
```

```
Alignment Scores:
Pred. No.: 0.0209 Length: 981
Score: 94.00 Matches: 18
Percent Similarity: 90.0% Conservative: 0
Best Local Similarity: 90.0% Mismatches: 2
Query Match: 83.2% Indels: 0
DB: 3 Gaps: 0
```

US-09-606-910E-1 (1-21) x US-10-029-517-16 (1-981)

```
QY 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db 39 GCCCATGGTGTACCTGGCCCGGACGACGAGCCGCTTGGGCTCCACCGCCCTCCA 98
```

Search completed: April 14, 2006, 19:10:39
Job time : 124.415 secs

GenCore version 5.1.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 14, 2006, 19:10:52 ; Search time 522.951 Seconds
(without alignments)
332.071 Million cell updates/sec

Title: US-09-606-910E-1
Perfect score: 113
Sequence: 1 AHGVTAPDTRPAGSTAPPA 21

Scoring table:
BLOSUM62
-Q/abss/ABSSWEB spool/US09606910/runat_14042006_091630_1777/app_query.fasta_1
-DB=Published_Applications_NA_Main -QMT=fastap -SUFFIX=p2n.rnpbm
-MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1
-MATRIX=blosum62 -TRANS=human4.0.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct
-THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMF=ptc -NORM=ext
-HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abes08
-USER=US09606910_@CGN 1_1364_@runat_14042006_091630_1777 -NCPU=6 -ICPU=3
-NO MMAP -NEG SCORES=0 -WAIT -DSBLOK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications_NA_Main:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	113	100.0	120	8	US-10-635-211-3
2	113	100.0	162	8	US-10-635-211-8
3	113	100.0	1424	7	US-10-447-839A-75
4	113	100.0	1424	8	US-10-778-859-75
5	113	100.0	1428	7	US-10-447-839A-20
6	113	100.0	1428	8	US-10-778-859-20
7	113	100.0	1527	5	US-10-057-136-19

8	113	100.0	1572	10	US-11-055-119-1
9	113	100.0	1721	3	US-09-864-864-280
10	113	100.0	1721	3	US-09-967-768A-224
11	113	100.0	1721	5	US-10-097-340-211
12	113	100.0	1721	5	US-10-171-311-155
13	113	100.0	1721	6	US-10-007-926A-58
14	113	100.0	1721	6	US-10-029-517-3
15	113	100.0	1721	6	US-10-172-118-775
16	113	100.0	1721	7	US-10-342-887-775
17	113	100.0	1721	8	US-10-775-920-88
18	113	100.0	1721	9	US-10-843-641A-6369
19	113	100.0	1721	9	US-10-498-704-3
20	113	100.0	1721	10	US-11-050-926-211
21	113	100.0	1761	8	US-10-635-211-1
22	113	100.0	1799	7	US-10-447-839A-19
23	113	100.0	1799	8	US-10-778-859-19
24	113	100.0	1804	3	US-09-964-824A-573
25	113	100.0	1804	6	US-10-029-517-17
26	113	100.0	1804	7	US-10-717-597-30
27	113	100.0	1804	8	US-10-775-920-84
28	113	100.0	1804	9	US-10-843-641A-5876
29	113	100.0	1804	9	US-10-756-149-1275
30	113	100.0	1804	9	US-10-498-704-17
31	113	100.0	1823	6	US-10-101-510-339
32	113	100.0	2026	5	US-10-198-846-12589
33	113	100.0	2238	8	US-10-775-920-87
34	113	100.0	2297	7	US-10-406-317-41
35	113	100.0	2678	6	US-10-252-157-103
36	113	100.0	4139	3	US-09-964-824A-105
37	113	100.0	4139	3	US-09-964-824A-578
38	113	100.0	4139	3	US-09-864-864-334
39	113	100.0	4139	3	US-09-880-107-2121
40	113	100.0	4139	3	US-09-968-007A-751
41	113	100.0	4139	5	US-10-171-311-157
42	113	100.0	4139	6	US-10-177-293-310
43	113	100.0	4139	7	US-10-440-464-155
44	113	100.0	4139	7	US-10-734-564-53
45	113	100.0	4139	8	US-10-775-920-80

ALIGNMENTS

RESULT 1

US-10-635-211-3
; Sequence 3, Application US/10635211
; Publication No. US20050031649A1
; GENERAL INFORMATION:
; APPLICANT: Beijing HYDAX Biotechnology Co. Ltd
; TITLE OF INVENTION: A recombinant fusion protein comprising BCG heat shock protein 6:
; TITLE OF INVENTION: and the epitope of MUC1
; FILE REFERENCE: FP03012US
; CURRENT APPLICATION NUMBER: US/10/635,211
; CURRENT FILING DATE: 2003-08-06
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 120
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(120)
US-10-635-211-3

Alignment Scores:	1.6e-06	Length:	120
Pred. No.:	113.00	Matches:	21
Score:	100.0%	Conservative:	0
Best Local Similarity:	100.0%	Mismatches:	0
Query Match:	100.0%	Indels:	0
DB:	8	Gaps:	0

US-09-606-910E-1 (1-21) x US-10-635-211-3 (1-120)

OTHER INFORMATION: Synthesized Sequence	
US-10-447-839A-75	
Alignment Scores:	
Pred. No.:	1.37e-05
Score:	113.00
Percent Similarity:	100.0%
Best Local Similarity:	100.0%
Query Match:	100.0%
DB:	7
US-09-606-910E-1 (1-21) x US-10-447-839A-75 (1-1424)	
Qy	1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db	1023 GCCACGGGTGTCACCTCGGCCCGGACACACGAGCGCGGCTCCACCGCCCCCA 964
Qy	21 Ala 21
Db	963 GCC 961
RESULT 4	
US-10-778-859-75/c	
; Sequence 75, Application US/10778859	
; Publication No. US20050042209A1	
; GENERAL INFORMATION:	
; APPLICANT: Kufe, Donald W.	
; APPLICANT: Ohno, Tsuneya	
; TITLE OF INVENTION: MUC1 EXTRACELLULAR DOMAIN AND CANCER TREATMENT COMPOSITIONS AND	
; TITLE OF INVENTION: METHODS DERIVED THEREFROM	
; FILE REFERENCE: 1000.05.011	
; CURRENT APPLICATION NUMBER: US/10/778,859	
; CURRENT FILING DATE: 2004-02-13	
; PRIOR APPLICATION NUMBER: 10/293,391	
; PRIOR FILING DATE: 2002-11-13	
; PRIOR APPLICATION NUMBER: 10/447,839	
; PRIOR FILING DATE: 2003-05-29	
; PRIOR APPLICATION NUMBER: 09/951,938	
; PRIOR FILING DATE: 2001-09-11	
; PRIOR APPLICATION NUMBER: 60/231,841	
; PRIOR FILING DATE: 2000-09-11	
; NUMBER OF SEQ ID NOS: 111	
; SOFTWARE: PatentIn version 3.2	
; SEQ ID NO 75	
; LENGTH: 1424	
; TYPE: RNA	
; ORGANISM: ARTIFICIAL	
; FEATURE:	
; OTHER INFORMATION: Synthesized Sequence	
US-10-778-859-75	
Alignment Scores:	
Pred. No.:	1.37e-05
Score:	113.00
Percent Similarity:	100.0%
Best Local Similarity:	100.0%
Query Match:	100.0%
DB:	8
US-09-606-910E-1 (1-21) x US-10-778-859-75 (1-1424)	
Qy	1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db	1023 GCCACGGGTGTCACCTCGGCCCGGACACACGAGCGCGGCTCCACCGCCCCCA 964
Qy	21 Ala 21
Db	963 GCC 961
RESULT 5	
US-10-447-839A-20	
; Sequence 20, Application US/10447839A	
; Publication No. US20040018181A1	

OTHER INFORMATION: Synthesized Sequence	
US-10-447-839A-75	
Alignment Scores:	
Pred. No.:	1.37e-05
Score:	113.00
Percent Similarity:	100.0%
Best Local Similarity:	100.0%
Query Match:	100.0%
DB:	7
US-09-606-910E-1 (1-21) x US-10-447-839A-75 (1-1424)	
Qy	1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db	1023 GCCACGGGTGTCACCTCGGCCCGGACACACGAGCGCGGCTCCACCGCCCCCA 964
Qy	21 Ala 21
Db	963 GCC 961
RESULT 4	
US-10-778-859-75/c	
; Sequence 75, Application US/10778859	
; Publication No. US20050042209A1	
; GENERAL INFORMATION:	
; APPLICANT: Kufe, Donald W.	
; APPLICANT: Ohno, Tsuneya	
; TITLE OF INVENTION: MUC1 EXTRACELLULAR DOMAIN AND CANCER TREATMENT COMPOSITIONS AND	
; TITLE OF INVENTION: METHODS DERIVED THEREFROM	
; FILE REFERENCE: 1000.05.011	
; CURRENT APPLICATION NUMBER: US/10/778,859	
; CURRENT FILING DATE: 2004-02-13	
; PRIOR APPLICATION NUMBER: 10/293,391	
; PRIOR FILING DATE: 2002-11-13	
; PRIOR APPLICATION NUMBER: 10/447,839	
; PRIOR FILING DATE: 2003-05-29	
; PRIOR APPLICATION NUMBER: 09/951,938	
; PRIOR FILING DATE: 2001-09-11	
; PRIOR APPLICATION NUMBER: 60/231,841	
; PRIOR FILING DATE: 2000-09-11	
; NUMBER OF SEQ ID NOS: 111	
; SOFTWARE: PatentIn version 3.2	
; SEQ ID NO 75	
; LENGTH: 1424	
; TYPE: RNA	
; ORGANISM: ARTIFICIAL	
; FEATURE:	
; OTHER INFORMATION: Synthesized Sequence	
US-10-778-859-75	
Alignment Scores:	
Pred. No.:	1.37e-05
Score:	113.00
Percent Similarity:	100.0%
Best Local Similarity:	100.0%
Query Match:	100.0%
DB:	8
US-09-606-910E-1 (1-21) x US-10-778-859-75 (1-1424)	
Qy	1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db	1023 GCCACGGGTGTCACCTCGGCCCGGACACACGAGCGCGGCTCCACCGCCCCCA 964
Qy	21 Ala 21
Db	963 GCC 961
RESULT 5	
US-10-447-839A-20	
; Sequence 20, Application US/10447839A	
; Publication No. US20040018181A1	


```

; GENERAL INFORMATION:
; APPLICANT: Kufe, Donald W.
; APPLICANT: Kharbanda, Surender
; APPLICANT: Weitman, Steven D.
; TITLE OF INVENTION: MUC1 INTERFERENCE RNA COMPOSITIONS AND METHODS DERIVED THEREFROM
; FILE REFERENCE: 1000.05.009
; CURRENT APPLICATION NUMBER: US/10/447,839A
; CURRENT FILING DATE: 2003-05-29
; PRIOR FILING DATE: 2002-11-13
; PRIOR FILING DATE: 2002-11-13
; PRIOR FILING DATE: 2003-05-29
; PRIOR FILING DATE: 2001-09-11
; PRIOR FILING DATE: 2001-09-11
; PRIOR FILING DATE: 2001-09-11
; PRIOR FILING DATE: 2000-09-11
; NUMBER OF SEQ ID NOS: 109
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 20
; LENGTH: 1428
; TYPE: RNA
; ORGANISM: RNA
US-10-447-839A-20

Alignment Scores:
Pred. No.: 1.37e-05 Length: 1428
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 7 Gaps: 0

US-09-606-910E-1 (1-21) x US-10-447-839A-20 (1-1428)

Qy 1 AlahisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db 403 GCCCAGGUGUACUCCUGGCCCGCCGACACACAGCGCGCGCGGCGUCCACCGCCCCCA 462

Qy 21 Ala 21
Db 463 GCC 465

RESULT 6
US-10-778-859-20
; Sequence 20, Application US/10778859
; Publication No. US20050042209A1
; GENERAL INFORMATION:
; APPLICANT: Kufe, Donald W.
; APPLICANT: Ohno, Tsuneya
; TITLE OF INVENTION: MUC1 EXTRACELLULAR DOMAIN AND CANCER TREATMENT COMPOSITIONS AND METHODS DERIVED THEREFROM
; FILE REFERENCE: 1000.05.011
; CURRENT APPLICATION NUMBER: US/10/778,859
; CURRENT FILING DATE: 2004-02-13
; PRIOR FILING DATE: 2002-11-13
; PRIOR FILING DATE: 2002-11-13
; PRIOR FILING DATE: 2003-05-29
; PRIOR FILING DATE: 2001-09-11
; PRIOR FILING DATE: 2001-09-11
; PRIOR FILING DATE: 2000-09-11
; NUMBER OF SEQ ID NOS: 111
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 20
; LENGTH: 1428
; TYPE: RNA
; ORGANISM: RNA
US-10-778-859-20

Alignment Scores:
Pred. No.: 1.37e-05 Length: 1428
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 7 Gaps: 0

US-09-606-910E-1 (1-21) x US-10-447-839A-20 (1-1428)

Qy 1 AlahisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db 403 GCCCAGGUGUACUCCUGGCCCGCCGACACACAGCGCGCGCGGCGUCCACCGCCCCCA 462

Qy 21 Ala 21
Db 463 GCC 465

RESULT 7
US-10-057-136-19
; Sequence 19, Application US/10057136
; Publication No. US20030021770A1
; GENERAL INFORMATION:
; APPLICANT: SCHLOM, JEFFREY
; APPLICANT: KANTOR, JUDITH
; APPLICANT: KUFU, DONALD
; APPLICANT: PANICALI, DENNIS
; APPLICANT: GRITZ, LINDA
; TITLE OF INVENTION: RECOMBINANT POX VIRUS FOR IMMUNIZATION AGAINST MUC1
; TITLE OF INVENTION: TUMOR-ASSOCIATED ANTIGEN
; FILE REFERENCE: 700953/47113C
; CURRENT APPLICATION NUMBER: US/10/057,136
; CURRENT FILING DATE: 2002-01-25
; PRIOR FILING DATE: 1999-08-03
; PRIOR FILING DATE: 1999-08-03
; PRIOR FILING DATE: 1998-02-24
; PRIOR FILING DATE: 1997-02-24
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
; LENGTH: 1527
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-057-136-19

Alignment Scores:
Pred. No.: 1.45e-05 Length: 1527
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 5 Gaps: 0

US-09-606-910E-1 (1-21) x US-10-057-136-19 (1-1527)

Qy 1 AlahisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db 244 GCCCAGGUGUACUCCUGGCCCGCCGACACACAGCGCGCGCGGCGUCCACCGCCCCCA 303

Qy 21 Ala 21
Db 304 GCC 306

RESULT 8
US-11-055-119-1
; Sequence 1, Application US/11055119
; Publication No. US20050142640A1
; GENERAL INFORMATION:
; APPLICANT: Taylor-Papadimitriou, Joyce
; APPLICANT: Heukamp, Lukas Carl
; APPLICANT: Offringa, Rienk
; APPLICANT: Melief, Cornelis Johanna Maria
; APPLICANT: Acres, Bruce
; APPLICANT: Thomas, Mireille
; TITLE OF INVENTION: MUC-1 derived peptides
; FILE REFERENCE: 029395-017
```

```

Query Match: 100.0% Indels: 0
DB: 8 Gaps: 0

US-09-606-910E-1 (1-21) x US-10-778-859-20 (1-1428)

Qy 1 AlahisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db 403 GCCCAGGUGUACUCCUGGCCCGCCGACACACAGCGCGCGCGGCGUCCACCGCCCCCA 462

Qy 21 Ala 21
Db 463 GCC 465
```



```
; APPLICANT: Karen LU
; APPLICANT: Rosemarie SCHMANDT
; APPLICANT: Xumei ZHAO
; APPLICANT: Karen GLATT
; TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
; FILE REFERENCE: MRI-030
; CURRENT APPLICATION NUMBER: US/10/097,340
; PRIOR FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: 60/276,025
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/325,149
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/276,026
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/324,967
; PRIOR FILING DATE: 2001/09/26
; PRIOR APPLICATION NUMBER: 60/311,732
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/325,102
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/323,580
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 211
; LENGTH: 1721
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-097-340-211

Alignment Scores:
Pred. No.: 1 61e-05 Length: 1721
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 5 Gaps: 0

US-09-606-910E-1 (1-21) x US-10-097-340-211 (1-1721)
Qy 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db 580 GCCACGGTGTCACTCGCGCCGACACACAGCGCGCGCGGCTCCACCGCCCCCA 639
Qy 21 Ala 21
Db 640 GCC 642

RESULT 12
US-10-171-311-155
; Sequence 155, Application US/10171311
; Publication NO. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoersh, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; PRIOR FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936

US-09-606-910E-1 (1-21) x US-10-097-340-211 (1-1721)
Qy 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db 580 GCCACGGTGTCACTCGCGCCGACACACAGCGCGCGGCTCCACCGCCCCCA 639
Qy 21 Ala 21
Db 640 GCC 642

RESULT 13
US-10-007-926A-58
; Sequence 58, Application US/10007926A
; Publication No. US20030143539A1
; GENERAL INFORMATION:
; APPLICANT: BERTUCCI, FRANCOIS
; APPLICANT: HOULGATTE, REMI
; APPLICANT: BIRNBAUM, DANIEL
; APPLICANT: NGUYEN, CATHERINE
; APPLICANT: VIENS, PATRICE
; APPLICANT: FERT, VINCENT
; TITLE OF INVENTION: GENE EXPRESSION PROFILING OF PRIMARY BREAST CARCINOMAS
; FILE REFERENCE: 1546-R-00
; CURRENT APPLICATION NUMBER: US/10/007,926A
; PRIOR FILING DATE: 2001-12-07
; PRIOR APPLICATION NUMBER: 60/254,090
; PRIOR FILING DATE: 2000-12-08
; NUMBER OF SEQ ID NOS: 468
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 58
; LENGTH: 1721
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: mucin 1, transmembrane (MUC1) gene.
US-10-007-926A-58

Alignment Scores:
Pred. No.: 1 61e-05 Length: 1721
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 6 Gaps: 0

US-09-606-910E-1 (1-21) x US-10-007-926A-58 (1-1721)
Qy 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db 580 GCCACGGTGTCACTCGCGCCGACACACAGCGCGCGGCTCCACCGCCCCCA 639
Qy 21 Ala 21
Db 640 GCC 642
```

```
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 155
; LENGTH: 1721
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-171-311-155

Alignment Scores:
Pred. No.: 1 61e-05 Length: 1721
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 5 Gaps: 0

US-09-606-910E-1 (1-21) x US-10-171-311-155 (1-1721)
Qy 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db 580 GCCACGGTGTCACTCGCGCCGACACACAGCGCGCGGCTCCACCGCCCCCA 639
Qy 21 Ala 21
Db 640 GCC 642

RESULT 13
US-10-007-926A-58
; Sequence 58, Application US/10007926A
; Publication No. US20030143539A1
; GENERAL INFORMATION:
; APPLICANT: BERTUCCI, FRANCOIS
; APPLICANT: HOULGATTE, REMI
; APPLICANT: BIRNBAUM, DANIEL
; APPLICANT: NGUYEN, CATHERINE
; APPLICANT: VIENS, PATRICE
; APPLICANT: FERT, VINCENT
; TITLE OF INVENTION: GENE EXPRESSION PROFILING OF PRIMARY BREAST CARCINOMAS
; FILE REFERENCE: 1546-R-00
; CURRENT APPLICATION NUMBER: US/10/007,926A
; PRIOR FILING DATE: 2001-12-07
; PRIOR APPLICATION NUMBER: 60/254,090
; PRIOR FILING DATE: 2000-12-08
; NUMBER OF SEQ ID NOS: 468
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 58
; LENGTH: 1721
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: mucin 1, transmembrane (MUC1) gene.
US-10-007-926A-58

Alignment Scores:
Pred. No.: 1 61e-05 Length: 1721
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 6 Gaps: 0

US-09-606-910E-1 (1-21) x US-10-007-926A-58 (1-1721)
Qy 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db 580 GCCACGGTGTCACTCGCGCCGACACACAGCGCGCGGCTCCACCGCCCCCA 639
Qy 21 Ala 21
Db 640 GCC 642
```

RESULT 14
US-10-029-517-3
; Sequence 3, Application US/10029517
; Publication No. US20030148969A1
; GENERAL INFORMATION:
; APPLICANT: Kenneth W. Dobie
; APPLICANT: Susan J. Myers
; TITLE OF INVENTION: ANTISENSE MODULATION OF MUCIN 1, TRANSMEMBRANE EXPRESSION
; FILE REFERENCE: RTS-0352
; CURRENT APPLICATION NUMBER: US/10/029,517
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 107
; SEQ ID NO 3
; LENGTH: 1721
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (58)...(1605)
US-10-029-517-3

Alignment Scores:
Pred. No.: 1,61e-05 Length: 1721
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 6 Gaps: 0

US-09-606-910E-1 (1-21) x US-10-029-517-3 (1-1721)

Qy 1 AlahisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db 580 GCCACGGTGTACCTCGGCCCGGACACAGCGCGCGGGGTCCACCGCCCCCA 639
Qy 21 Ala 21
Db 640 GCC 642

RESULT 15
US-10-172-118-775
; Sequence 775, Application US/10172118
; Publication No. US20030224374A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Chris
; APPLICANT: Van 't Veer, Laura
; APPLICANT: Van de Vijver, Marc
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-175-999
; CURRENT APPLICATION NUMBER: US/10/172,118
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/380,770
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 775
; LENGTH: 1721
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NM_002456
; DATABASE ENTRY DATE: 2001-06-18
US-10-172-118-775

Alignment Scores:
Pred. No.: 1,61e-05 Length: 1721
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0

Query Match: 100.0% Indels: 0
DB: 6 Gaps: 0
US-09-606-910E-1 (1-21) x US-10-172-118-775 (1-1721)
Qy 1 AlahisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db 580 GCCACGGTGTACCTCGGCCCGGACACAGCGCGCGGGGTCCACCGCCCCCA 639
Qy 21 Ala 21
Db 640 GCC 642

Search completed: April 14, 2006, 19:35:18
Job time : 523.951 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 14, 2006, 19:18:21 ; Search time 352.902 Seconds
(without alignments)
239.682 Million cell updates/sec

Title: US-09-606-910E-1
Perfect score: 113
Sequence: 1 AHGVTAPDTRPAPGSTAPPA 21

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 9281099 seqs, 2013915447 residues

Total number of hits satisfying chosen parameters: 18562198

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+p2n.model -DEV=xlh
-Q=/abss/ABSSWEB_spool/US09606910/runat_14042006_091631_1823/app_query.fasta.1
-DB=Published Applications NA New -QFMT=fastap -SUFFIX=p2n.rnpbn -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=2000000000 -HOST=abss08
-USER=US09606910 @CGN_1_1_975 @runat_14042006_091631_1823 -NCPU=6 -ICPU=3
-NO MMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA_New.*

1: /SIDSS/ptodata/2/pubpna/US08_NEW_PUB.seq.*
2: /SIDSS/ptodata/2/pubpna/US06_NEW_PUB.seq.*
3: /SIDSS/ptodata/2/pubpna/US07_NEW_PUB.seq.*
4: /SIDSS/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
5: /SIDSS/ptodata/2/pubpna/US05_NEW_PUB.seq.*
6: /SIDSS/ptodata/2/pubpna/US09_NEW_PUB.seq.*
7: /SIDSS/ptodata/2/pubpna/US10_NEW_PUB.seq.*
8: /SIDSS/ptodata/2/pubpna/US10_NEW_PUB.seq.*
9: /SIDSS/ptodata/2/pubpna/US10_NEW_PUB.seq.*
10: /SIDSS/ptodata/2/pubpna/US10_NEW_PUB.seq.*
11: /SIDSS/ptodata/2/pubpna/US11_NEW_PUB.seq.*
12: /SIDSS/ptodata/2/pubpna/US11_NEW_PUB.seq.*
13: /SIDSS/ptodata/2/pubpna/US11_NEW_PUB.seq.*
14: /SIDSS/ptodata/2/pubpna/US11_NEW_PUB.seq.*
15: /SIDSS/ptodata/2/pubpna/US60_NEW_PUB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	113	100.0	328	9	US-10-517-696-41 Sequence 41, Appl
2	113	100.0	1194	9	US-10-517-696-83 Sequence 83, Appl

3	113	100.0	1378	9	US-10-517-696-64	Sequence 64, Appl
4	113	100.0	1630	9	US-10-517-696-79	Sequence 79, Appl
5	113	100.0	1634	9	US-10-517-696-60	Sequence 60, Appl
6	113	100.0	1712	9	US-10-517-696-57	Sequence 57, Appl
7	113	100.0	1721	11	US-11-233-510-19	Sequence 19, Appl
8	113	100.0	1738	9	US-10-517-696-40	Sequence 40, Appl
9	113	100.0	1755	9	US-10-517-696-44	Sequence 44, Appl
10	113	100.0	1774	7	US-10-515-872-9	Sequence 9, Appl
11	113	100.0	1803	9	US-10-517-696-70	Sequence 70, Appl
12	113	100.0	1804	9	US-10-501-035-113	Sequence 113, Appl
13	113	100.0	1808	9	US-10-517-696-77	Sequence 77, Appl
14	113	100.0	1823	9	US-10-517-696-78	Sequence 78, Appl
15	113	100.0	1874	9	US-10-517-696-59	Sequence 59, Appl
16	113	100.0	1882	9	US-10-517-696-48	Sequence 48, Appl
17	113	100.0	1918	9	US-10-517-696-43	Sequence 43, Appl
18	113	100.0	1930	9	US-10-517-696-49	Sequence 49, Appl
19	113	100.0	1945	9	US-10-517-696-47	Sequence 47, Appl
20	113	100.0	1949	9	US-10-517-696-69	Sequence 69, Appl
21	113	100.0	1953	9	US-10-517-696-39	Sequence 39, Appl
22	113	100.0	2045	9	US-10-517-696-72	Sequence 72, Appl
23	113	100.0	2049	9	US-10-517-696-53	Sequence 53, Appl
24	113	100.0	2090	9	US-10-517-696-76	Sequence 76, Appl
25	113	100.0	2094	9	US-10-517-696-52	Sequence 52, Appl
26	113	100.0	2194	9	US-10-517-696-54	Sequence 54, Appl
27	113	100.0	2255	9	US-10-517-696-38	Sequence 38, Appl
28	113	100.0	2333	9	US-10-517-696-56	Sequence 56, Appl
29	113	100.0	4139	9	US-10-826-585-39	Sequence 39, Appl
30	113	100.0	4144	9	US-10-517-696-37	Sequence 37, Appl
31	113	100.0	6245	8	US-10-401-386B-61	Sequence 61, Appl
32	98	86.7	468	8	US-10-401-386B-43	Sequence 43, Appl
33	98	86.7	921	8	US-10-401-386B-30	Sequence 30, Appl
34	98	86.7	3343	8	US-10-401-386B-32	Sequence 32, Appl
35	94	83.2	1404	7	US-10-515-872-13	Sequence 13, Appl
36	87	77.0	1030	9	US-10-517-696-42	Sequence 42, Appl
37	87	77.0	1605	9	US-10-517-696-58	Sequence 58, Appl
38	87	77.0	1798	9	US-10-517-696-50	Sequence 50, Appl
39	62	54.9	918	14	US-11-165-226-123	Sequence 123, Appl
40	60	53.1	431	6	US-09-925-065A-816944	Sequence 816944,
41	60	53.1	436	6	US-09-925-065A-857612	Sequence 857612,
42	60	53.1	1878	6	US-09-925-065A-68959	Sequence 68959, A
43	60	53.1	1878	6	US-09-925-065A-68961	Sequence 68961, A
44	60	53.1	1878	9	US-10-301-480-170198	Sequence 170198,
45	60	53.1	1878	9	US-10-301-480-170200	Sequence 170200,

ALIGNMENTS

RESULT 1

US-10-517-696-41
; Sequence 41, Application US/10517696
; Publication No. US20060051759A1
; GENERAL INFORMATION:
; APPLICANT: diadexus, Inc.
; APPLICANT: Saiceda, Susana
; APPLICANT: Macina, Roberto A.
; APPLICANT: Turner, Leah R.
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and P.
; FILE REFERENCE: DEX-0432
; CURRENT APPLICATION NUMBER: US/10/517,696
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 60/389,327
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 171
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 41
; LENGTH: 328
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-517-696-41

Alignment Scores:

```

Pred. No.: 0.000138 Length: 328
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 9 Gaps: 0

US-09-606-910E-1 (1-21) x US-10-517-696-41 (1-328)

Qy 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db 113 GCCACGGTGTCACTCGGCCCGGACACACAGCGCGCGGCTCCACCGCCCCCA 172

Qy 21 Ala 21
Db 173 GCC 175

RESULT 2
US-10-517-696-83
; Sequence 83, Application US/10517696
; Publication No. US20060051759A1
; GENERAL INFORMATION:
; APPLICANT: diadexus, Inc.
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto A.
; APPLICANT: Turner, Leah R.
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and Pr
; FILE REFERENCE: DEX-0432
; CURRENT APPLICATION NUMBER: US/10/517,696
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 60/389,327
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 171
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 83
; LENGTH: 1194
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-517-696-83

Alignment Scores:
Pred. No.: 0.000407 Length: 1194
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 9 Gaps: 0

US-09-606-910E-1 (1-21) x US-10-517-696-83 (1-1194)

Qy 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db 614 GCCACGGTGTCACTCGGCCCGGACACACAGCGCGCGGCTCCACCGCCCCCA 673

Qy 21 Ala 21
Db 674 GCC 676

RESULT 3
US-10-517-696-64
; Sequence 64, Application US/10517696
; Publication No. US20060051759A1
; GENERAL INFORMATION:
; APPLICANT: diadexus, Inc.
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto A.
; APPLICANT: Turner, Leah R.
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and Pr
; FILE REFERENCE: DEX-0432
```

```

; CURRENT APPLICATION NUMBER: US/10/517,696
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 60/389,327
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 171
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 64
; LENGTH: 1378
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-517-696-64

Alignment Scores:
Pred. No.: 0.000459 Length: 1378
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 9 Gaps: 0

US-09-606-910E-1 (1-21) x US-10-517-696-64 (1-1378)

Qy 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db 614 GCCACGGTGTCACTCGGCCCGGACACACAGCGCGCGGCTCCACCGCCCCCA 673

Qy 21 Ala 21
Db 674 GCC 676

RESULT 4
US-10-517-696-79
; Sequence 79, Application US/10517696
; Publication No. US20060051759A1
; GENERAL INFORMATION:
; APPLICANT: diadexus, Inc.
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto A.
; APPLICANT: Turner, Leah R.
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and Pr
; FILE REFERENCE: DEX-0432
; CURRENT APPLICATION NUMBER: US/10/517,696
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 60/389,327
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 171
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 79
; LENGTH: 1630
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-517-696-79

Alignment Scores:
Pred. No.: 0.000529 Length: 1630
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 9 Gaps: 0

US-09-606-910E-1 (1-21) x US-10-517-696-79 (1-1630)

Qy 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db 614 GCCACGGTGTCACTCGGCCCGGACACACAGCGCGCGGCTCCACCGCCCCCA 673

Qy 21 Ala 21
Db 674 GCC 676
```

```
RESULT 5
US-10-517-696-60
; Sequence 60, Application US/10517696
; Publication No. US20060051759A1
; GENERAL INFORMATION:
; APPLICANT: diadexus, Inc.
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto A.
; APPLICANT: Turner, Leah R.
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and P
; FILE REFERENCE: DEX-0432
; CURRENT APPLICATION NUMBER: US/10/517,696
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 60/389,327
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 171
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 60
; LENGTH: 1634
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-517-696-60

Alignment Scores:
Pred. No.: 0.00053      Length: 1634
Score: 113.00          Matches: 21
Percent Similarity: 100.0%      Conservative: 0
Best Local Similarity: 100.0%      Mismatches: 0
Query Match: 100.0%      Indels: 0
DB: 9                  Gaps: 0

US-09-606-910E-1 (1-21) x US-10-517-696-60 (1-1634)

Qy 1 AlahisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db 614 GCCCACGGTGTCACTCGGCCCGGACACCGAGCGCGCGGCTCCACCGCCCCCA 673

Qy 21 Ala 21
Db 674 GCC 676

RESULT 6
US-10-517-696-57
; Sequence 57, Application US/10517696
; Publication No. US20060051759A1
; GENERAL INFORMATION:
; APPLICANT: diadexus, Inc.
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto A.
; APPLICANT: Turner, Leah R.
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and P
; FILE REFERENCE: DEX-0432
; CURRENT APPLICATION NUMBER: US/10/517,696
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 60/389,327
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 171
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 57
; LENGTH: 1712
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-517-696-57

Alignment Scores:
Pred. No.: 0.000551     Length: 1712
Score: 113.00          Matches: 21
Percent Similarity: 100.0%      Conservative: 0
Best Local Similarity: 100.0%      Mismatches: 0
```

```
Query Match: 100.0%      Indels: 0
DB: 9                  Gaps: 0

US-09-606-910E-1 (1-21) x US-10-517-696-57 (1-1712)

Qy 1 AlahisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db 614 GCCCACGGTGTCACTCGGCCCGGACACCGAGCGCGGCTCCACCGCCCCCA 673

Qy 21 Ala 21
Db 674 GCC 676

RESULT 7
US-11-233-510-19
; Sequence 19, Application US/11233510
; Publication No. US20060063190A1
; GENERAL INFORMATION:
; APPLICANT: Fischer, Timothy J.
; APPLICANT: Whitehead, Clark M.
; APPLICANT: Malinowski, Douglas P.
; APPLICANT: Marcelpoil, Raphael
; APPLICANT: Morel, Didier
; TITLE OF INVENTION: Methods and Compositions for Evaluating
; TITLE OF INVENTION: Breast Cancer Prognosis
; FILE REFERENCE: 46143/296738
; CURRENT APPLICATION NUMBER: US/11/233,510
; CURRENT FILING DATE: 2005-09-22
; PRIOR APPLICATION NUMBER: 60/612,073
; PRIOR FILING DATE: 2004-09-22
; PRIOR APPLICATION NUMBER: 60/611,965
; PRIOR FILING DATE: 2004-09-22
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 1721
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (58)...(1605)
US-11-233-510-19

Alignment Scores:
Pred. No.: 0.000554     Length: 1721
Score: 113.00          Matches: 21
Percent Similarity: 100.0%      Conservative: 0
Best Local Similarity: 100.0%      Mismatches: 0
Query Match: 100.0%      Indels: 0
DB: 11                 Gaps: 0

US-09-606-910E-1 (1-21) x US-11-233-510-19 (1-1721)

Qy 1 AlahisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db 580 GCCCACGGTGTCACTCGGCCCGGACACCGAGCGCGGCTCCACCGCCCCCA 639

Qy 21 Ala 21
Db 640 GCC 642

RESULT 8
US-10-517-696-40
; Sequence 40, Application US/10517696
; Publication No. US20060051759A1
; GENERAL INFORMATION:
; APPLICANT: diadexus, Inc.
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto A.
; APPLICANT: Turner, Leah R.
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and P
```

```

; FILE REFERENCE: DEX-0432
; CURRENT APPLICATION NUMBER: US/10/517,696
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 60/389,327
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 171
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 40
; LENGTH: 1738
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-517-696-40

Alignment Scores:
Pred. No.: 0.000558 Length: 1738
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 9 Gaps: 0

US-09-606-910E-1 (1-21) x US-10-517-696-40 (1-1738)

Qy 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db 476 GCCCACGGTGTACCTCGCGCCCGGACACACAGCGCGCGGGCTCCACCGCCCCCA 535
Qy 21 Ala 21
Db 536 GCC 538

RESULT 9
US-10-517-696-44
; Sequence 44, Application US/10517696
; Publication No. US20060051759A1
; GENERAL INFORMATION:
; APPLICANT: diadexus, Inc.
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto A.
; APPLICANT: Turner, Leah R.
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and P
; FILE REFERENCE: DEX-0432
; CURRENT APPLICATION NUMBER: US/10/517,696
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 60/389,327
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 171
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 44
; LENGTH: 1755
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1682)..(1682)
; OTHER INFORMATION: n=a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1733)..(1733)
; OTHER INFORMATION: n=a, c, g, or t
US-10-517-696-44

Alignment Scores:
Pred. No.: 0.000563 Length: 1755
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 9 Gaps: 0

US-09-606-910E-1 (1-21) x US-10-517-696-44 (1-1755)

```

```

Qy 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db 587 GCCCACGGTGTACCTCGCGCCCGGACACACAGCGCGCGGGCTCCACCGCCCCCA 646
Qy 21 Ala 21
Db 647 GCC 649

RESULT 10
US-10-515-872-9
; Sequence 9, Application US/10515872
; Publication No. US20060062798A1
; GENERAL INFORMATION:
; APPLICANT: BURDEN, NEIL
; APPLICANT: HAMBLIN, PAUL
; TITLE OF INVENTION: VACCINES
; FILE REFERENCE: PG4852
; CURRENT APPLICATION NUMBER: US/10/515,872
; CURRENT FILING DATE: 2004-11-24
; PRIOR APPLICATION NUMBER: PCT/EP03/05595
; PRIOR FILING DATE: 2003-05-23
; PRIOR APPLICATION NUMBER: GB 0212036.8
; PRIOR FILING DATE: 2002-05-24
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1774
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-515-872-9

Alignment Scores:
Pred. No.: 0.000568 Length: 1774
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 7 Gaps: 0

US-09-606-910E-1 (1-21) x US-10-515-872-9 (1-1774)

Qy 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db 480 GCCCACGGTGTACCTCGCGCCCGGACACACAGCGCGCGGGCTCCACCGCCCCCA 539
Qy 21 Ala 21
Db 540 GCC 542

RESULT 11
US-10-517-696-70
; Sequence 70, Application US/10517696
; Publication No. US20060051759A1
; GENERAL INFORMATION:
; APPLICANT: diadexus, Inc.
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto A.
; APPLICANT: Turner, Leah R.
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and P
; FILE REFERENCE: DEX-0432
; CURRENT APPLICATION NUMBER: US/10/517,696
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 60/389,327
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 171
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 70
; LENGTH: 1803
; TYPE: DNA
; ORGANISM: Homo sapien

```



```
US-10-517-696-70
Alignment Scores:
Pred. No.: 0.000576 Length: 1803
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 9 Gaps: 0

US-09-606-910E-1 (1-21) x US-10-517-696-70 (1-1803)
QY 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db 472 GCCACGGTGTACCTCGCGCCGGACACAGCGCGCGCGCGGTCCACCGCCCCCA 531
QY 21 Ala 21
Db 532 GCC 534

RESULT 12
US-10-501-035-113
; Sequence 113, Application US/10501035
; Publication No. US20060046249A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: IDENTIFICATION OF POLYNUCLEOTIDES AND POLYPEPTIDE FOR PREDICTING
; TITLE OF INVENTION: ACTIVITY OF COMPOUNDS THAT INTERACT WITH PROTEIN TYROSINE KINASE
; TITLE OF INVENTION: AND/OR PROTEIN TYROSINE KINASE PATHWAYS
; FILE REFERENCE: D0185 PCT
; CURRENT APPLICATION NUMBER: US/10/501,035
; CURRENT FILING DATE: 2004-07-09
; PRIOR APPLICATION NUMBER: US 60/350,061
; PRIOR FILING DATE: 2002-01-18
; NUMBER OF SEQ ID NOS: 795
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 113
; LENGTH: 1804
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-501-035-113
Alignment Scores:
Pred. No.: 0.000576 Length: 1804
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 9 Gaps: 0

US-09-606-910E-1 (1-21) x US-10-501-035-113 (1-1804)
QY 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db 475 GCCACGGTGTACCTCGCGCCGGACACAGCGCGCGCGGTCCACCGCCCCCA 534
QY 21 Ala 21
Db 535 GCC 537

RESULT 13
US-10-517-696-77
; Sequence 77, Application US/10517696
; Publication No. US20060051759A1
; GENERAL INFORMATION:
; APPLICANT: diaDexus, Inc.
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto A.
; APPLICANT: Turner, Leah R.
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and Pr
; FILE REFERENCE: DEX-0432
```

```
; CURRENT APPLICATION NUMBER: US/10/517,696
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 60/389,327
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 171
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 77
; LENGTH: 1808
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-517-696-77
Alignment Scores:
Pred. No.: 0.000577 Length: 1808
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 9 Gaps: 0

US-09-606-910E-1 (1-21) x US-10-517-696-77 (1-1808)
QY 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db 614 GCCACGGTGTACCTCGCGCCGGACACAGCGCGCGCGGTCCACCGCCCCCA 673
QY 21 Ala 21
Db 674 GCC 676

RESULT 14
US-10-517-696-78
; Sequence 78, Application US/10517696
; Publication No. US20060051759A1
; GENERAL INFORMATION:
; APPLICANT: diaDexus, Inc.
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto A.
; APPLICANT: Turner, Leah R.
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and P
; FILE REFERENCE: DEX-0432
; CURRENT APPLICATION NUMBER: US/10/517,696
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 60/389,327
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 171
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 78
; LENGTH: 1823
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-517-696-78
Alignment Scores:
Pred. No.: 0.000581 Length: 1823
Score: 113.00 Matches: 21
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 9 Gaps: 0

US-09-606-910E-1 (1-21) x US-10-517-696-78 (1-1823)
QY 1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db 614 GCCACGGTGTACCTCGCGCCGGACACAGCGCGCGCGGTCCACCGCCCCCA 673
QY 21 Ala 21
Db 674 GCC 676
```

```
RESULT 15
US-10-517-696-59
; Sequence 59, Application US/10517696
; Publication No. US20060051759A1
; GENERAL INFORMATION:
; APPLICANT: diadexus, Inc.
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto A.
; APPLICANT: Turner, Leah R.
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and Pr
; FILE REFERENCE: DEX-0432
; CURRENT APPLICATION NUMBER: US/10/517,696
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 60/389,327
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 171
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 59
; LENGTH: 1874
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-517-696-59

Alignment Scores:
Pred. No.:      0.000595      Length:      1874
Score:          113.00      Matches:      21
Percent Similarity: 100.0%      Conservative: 0
Best Local Similarity: 100.0%      Mismatches: 0
Query Match:      100.0%      Indels:      0
DB:               9      Gaps:      0

US-09-606-910E-1 (1-21) x US-10-517-696-59 (1-1874)

QY      1 AlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 20
Db      614 GCCACGGTGTCACTCGGCCCGCCGACACAGGCCGCCCGCGCTCCACGCCCGCCCA 673

QY      21 Ala 21
Db      674 GCC 676
```

Search completed: April 14, 2006, 19:46:53
Job time : 353.902 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 14, 2006, 18:40:06 ; Search time 368.78 Seconds

(without alignments)
361.445 Million cell updates/sec

Title: US-09-606-910E-2

Perfect score: 109

Sequence: 1 HGVTSAPDTRPAGSTAPPA 20

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4996997 seqs, 332346308 residues

Total number of hits satisfying chosen parameters: 993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/abss/ABSSWEB/spool/US09606910/runat_14042006_091618_1498/app_query.fasta 1
-DB=N Geneseq -QFMT=faetap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-LOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abss08
-USER=US09606910 @CNC 1 1 900 @runat_14042006_091618_1498 -NCPU=6 -ICPU=3
-NO MMAP -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N Geneseq_21.*
1: Geneseqn1980s.*
2: Geneseqn1990s.*
3: Geneseqn2000s.*
4: Geneseqn2001as.*
5: Geneseqn2001bs.*
6: Geneseqn2002as.*
7: Geneseqn2002bs.*
8: Geneseqn2003as.*
9: Geneseqn2003bs.*
10: Geneseqn2003cs.*
11: Geneseqn2003ds.*
12: Geneseqn2004as.*
13: Geneseqn2004bs.*
14: Geneseqn2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	109	100.0	120	14	ADx05274 Human muc
C 2	109	100.0	156	10	ADk68635 HSP65-MUC
C 3	109	100.0	162	14	ADx05279 HSP65-MUC
C 4	109	100.0	309	1	AAn90579 pDF9.3 cD

5	109	100.0	328	12	ADi57670
6	109	100.0	525	3	AA000385
7	109	100.0	891	3	AA000391
8	109	100.0	1125	14	ADz87487 Human MUC
9	109	100.0	1194	12	ADi57712 Human bre
10	109	100.0	1371	3	AA000388 Human MUC
11	109	100.0	1378	12	ADi57693
12	109	100.0	1424	12	AD023180
13	109	100.0	1428	6	ABL60159
14	109	100.0	1428	12	AD023125
15	109	100.0	1457	12	ADF32627
16	109	100.0	1527	2	AAV48329
17	109	100.0	1548	14	AEA11045
18	109	100.0	1548	14	AEA12738
19	109	100.0	1548	14	AE000289
20	109	100.0	1572	5	AA000585
21	109	100.0	1614	12	ADK70370
22	109	100.0	1630	12	ADi57708
23	109	100.0	1634	12	ADi57689
24	109	100.0	1712	12	ADi57686
25	109	100.0	1713	13	ADW78695
26	109	100.0	1721	6	ABS76475
27	109	100.0	1721	6	ABL68032
28	109	100.0	1721	6	ABK09743
29	109	100.0	1721	6	ABV94067
30	109	100.0	1721	8	ADA50566
31	109	100.0	1721	8	ACF12906
32	109	100.0	1721	9	AA056938
33	109	100.0	1721	12	ADK65926
34	109	100.0	1721	13	ADR24914
35	109	100.0	1737	3	AA000394
36	109	100.0	1738	12	ADi57669
37	109	100.0	1751	12	ADi57673
38	109	100.0	1761	14	ADX05272
39	109	100.0	1774	12	ADe43991
40	109	100.0	1774	12	ADP33625
41	109	100.0	1799	12	AD023124
42	109	100.0	1800	13	ADR89853
43	109	100.0	1803	12	ADi57699
44	109	100.0	1804	6	ABL67539
45	109	100.0	1804	9	AA056950

ALIGNMENTS

RESULT 1
ADx05274
ID ADX05274 standard; DNA; 120 BP.
XX
AC ADX05274;
XX
DT 21-APR-2005 (first entry)
XX
DE Human mucin MUC1 coding sequence, SEQ ID 3.
XX
KW Fusion protein; Cytostatic; Gene Therapy; mucin; carcinoma; gene;
KW ds; chromosome 1.
XX
OS Homo sapiens.
XX
FH Key
CDS Location/Qualifiers
FT /*tag= a
FT /product= "MUC1 epitope"
FT /partial
FT /note= "No start or stop codon"
XX
PN US2005031649-A1.
XX
PD 10-FEB-2005.
XX
PF 06-AUG-2003; 2003US-00635211.
XX

CC ME). The BCG HSP65 is located at the amino terminal portion of the fusion
CC protein and the epitope of MUC1 is located at the carboxy terminal
CC portion of the fusion protein. The MUC1 epitope can generate MUC1
CC specific cytotoxic T lymphocytes. The recombinant fusion protein is
CC useful for treating or preventing human carcinomas, preferably MUC1
CC expressing carcinomas. The present sequence was used in an example for
CC producing the fusion protein of the invention.

XX SQ Sequence 162 BP; 39 A; 43 C; 58 G; 22 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 0.000183 Length: 162
Score: 109.00 Matches: 20
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 14 Gaps: 0

US-09-606-910E-2 (1-20) x ADX05279 (1-162)

Qy 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20

Db 117 CACGGTGTACTCTGCTCCGGACACCGTCCGGTCCGGGTCTACCGCTCCGGCGCT 58

RESULT 4

AD AN90579/C
ID AN90579 standard; cDNA; 309 BP.

XX AC AN90579;

XX XX 27-AUG-2003 (revised)

DT 25-MAR-2003 (revised)

DT 04-DEC-1989 (first entry)

XX pDF9.3 cDNA insert.

DE pDF9.3; human DF3 breast carcinoma-associated antigen epitope.

XX Human MCF-7 breast carcinoma cells; 'ATCC HTB22'.

XX WO8907107-A.

XX 10-AUG-1989.

XX 29-JAN-1988; 88US-00149831.

XX 29-JAN-1988; 88US-00149831.

XX (DANA-) DANA-FARBER CANCER.

XX Kufe DW;

XX WPI; 1989-248989/34.

DR P-PSDB; AAP91045, AAP91053, AAP91054, AAP90146.

XX Recombinant polypeptide(s) - contains DF 3 breast carcinoma antigen
PT epitope and useful as assay reagents, and encoding DNA sequences.

XX Claim 1; Fig 4; 31pp; English.

XX The sequence encodes a carbohydrate-free polypeptide contg. DF3 breast
CC carcinoma antigen epitope. Useful as a competitive binding assay reagent
CC and improves diagnosis. The cDNA consists of nearly identical 60 BP
CC tandem repeats which are 85% GC-rich. See also AAP91045, AAP91046,
CC AAP91053, AAP91054. (Updated on 25-MAR-2003 to correct PA field.)
CC (Updated on 27-AUG-2003 to correct OS field.)

XX SQ Sequence 309 BP; 21 A; 92 C; 160 G; 36 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 0.000352 Length: 309
Score: 109.00 Matches: 20
Percent Similarity: 100.0% Conservative: 0

Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 1 Gaps: 0

US-09-606-910E-2 (1-20) x AAN90579 (1-309)

Qy 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20

Db 305 CACGGTGTACCTCGGCCCGGACACCGAGCGCGGCTCCACCGCCCCCAGCC 246

RESULT 5

AD I57670
ID ADI57670 standard; cDNA; 328 BP.

XX AC ADI57670;

XX DT 22-APR-2004 (first entry)

XX DE Human breast specific nucleic acid (BSNA) #41.

XX Human; breast specific nucleic acid; BSNA; gene; ss; metastasis;
KW breast cancer; cytostatic.

XX OS Homo sapiens.

XX FN WO2003106648-A2.

XX PD 24-DEC-2003.

XX PF 16-JUN-2003; 2003WO-US018934.

XX PR 14-JUN-2002; 2002US-0389327P.

XX PA (DIAD-) DIADEXUS INC.

XX PI Salceda S, Macina RA, Turner LR, Sun Y, Liu C;

XX WPI; 2004-082185/08.

DR P-PSDB; ADI57744.

XX Novel isolated polypeptide comprising breast specific protein sequences,
PT useful for diagnosing or monitoring presence and metastases of breast
PT cancer in patient.

XX Claim 1; SEQ ID NO 41; 370pp; English.

XX The invention relates to human breast specific nucleic acids (BSNA) and
CC the breast specific proteins (BSP) they encode. The nucleic acids are
CC useful for determining the presence of a BSNA in a sample which involves
CC contacting the sample with a BSNA under conditions in which the BSNA will
CC selectively hybridise to a BSNA in the sample, and detecting the
CC hybridisation. The nucleic acids are useful for determining the presence
CC of a BSP in a sample which involves contacting the sample with suitable
CC reagent under conditions in which the reagent will selectively interact
CC with the BSP, and detecting the interaction of the reagent with a BSP in
CC the sample. The nucleic acids and proteins are useful for diagnosing or
CC monitoring the presence and metastases of breast cancer in a patient,
CC which involves determining an amount of nucleic acid or protein and
CC comparing the determined amount of nucleic acid or protein in the sample
CC of the patient to the amount of a breast specific marker in a normal
CC control, where a difference in the determined amount in the sample
CC compared to the amount in the control is associated with the presence of
CC breast cancer. The sequences are useful for treating a patient with
CC breast cancer, involving administering a composition consisting of a BSNA
CC or a BSP to a patient, where the administration induces an immune
CC response against the breast cancer cell expressing the BSNA or BSP. This
CC sequence represents a human BSNA of the invention.

XX SQ Sequence 328 BP; 47 A; 124 C; 95 G; 61 T; 0 U; 1 Other;

Alignment Scores:
Pred. No.: 0.000374 Length: 328
Score: 109.00 Matches: 20

Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 12 Gaps: 0

US-09-606-910E-2 (1-20) x ADI57670 (1-328)

Qy 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
Db 116 CACGGTGTACCTCGGCCCGGACACCCAGGCGCGCGGCTCCACCGCCCCCAGCC 175

RESULT 6

AAD00385

ID AAD00385 standard; DNA; 525 BP.

AC AAD00385;

DT 29-AUG-2000 (first entry)

XX Human Mucin 1 (MUC-1) protein fragment encoding DNA #2.

XX Human; Mucin 1; tumour; pMRS30 expression vector; anti-tumour;

XX therapy; immune response; cytostatic; vaccine; ds.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 1..525

XX /tag= a

XX /product= "MUC-1 protein fragment"

XX WO200025827-A2.

XX 11-MAY-2000.

XX 18-OCT-1999; 99WO-EP007874.

XX 30-OCT-1998; 98IT-MI002330.

XX (MENA) MENARINI RICERCHE SPA.

XX Parente D, Di Massimo AM, De Santis R;

XX WPI; 2000-365410/31.

XX P-PSDB; AAY71021.

XX Composition containing one or more DNA molecules encoding fragments of a

XX Mucin 1 (MUC-1) protein overexpressed in tumor cells, useful in anti-

XX tumor therapy.

XX Claim 16; Fig 2; 56pp; English.

XX The present sequence is a DNA encoding a fragment of human Mucin 1 (MUC-

XX 1) antigenic protein which is overexpressed in tumour cells. The sequence

XX was obtained from BT20 tumour cells by reverse transcriptase-PCR and

XX corresponds to nucleotides 205-720 of the EMBL sequence J05581 with a

XX start codon and two stop codons. The present sequence is cloned into a

XX pMRS30 expression vector and used in pharmaceutical composition e.g.

XX vaccine for inducing an antigen-specific anti-tumour immune response.

XX Composition containing this DNA molecule is useful in anti-tumour therapy

XX of patients affected with tumours characterised by high MUC-1 expression

XX SQ Sequence 525 BP; 111 A; 199 C; 128 G; 87 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 0.000603 Length: 525

Score: 109.00 Matches: 20

Percent Similarity: 100.0% Conservative: 0

Best Local Similarity: 100.0% Mismatches: 0

Query Match: 100.0% Indels: 0

DB: 3 Gaps: 0

US-09-606-910E-2 (1-20) x AAD00385 (1-525)

Qy 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
Db 277 CACGGTGTACCTCGGCTCCGATACACAGGCGCGCGGCTAGTACCGCCCTCTGCC 336

RESULT 7

AAD00391

ID AAD00391 standard; DNA; 891 BP.

AC AAD00391;

XX 15-SEP-2003 (revised)

XX 29-AUG-2000 (first entry)

XX Ubiquitin-E. coli LacI-human Mucin 1 fusion protein encoding DNA #2.

XX Ubiquitin; LacI; beta-galactosidase; fusion protein; human; Mucin 1;

XX MUC-1; tumour; pMRS30 expression vector; anti-tumour; therapy;

XX immune response; cytostatic; vaccine; ds.

XX Homo sapiens.

XX Escherichia coli.

XX Chimeric.

XX Key Location/Qualifiers

XX CDS 1..891

XX /tag= a

XX /product= "Ubiquitin-LacI-MUC-1 fusion protein"

XX misc_feature 1..369

XX /tag= b

XX /label= UBILacI_DNA

XX /note= "Includes ubiquitin-E. coli LacI fusion DNA"

XX misc_feature 370..891

XX /tag= c

XX /note= "Human MUC-1 partial DNA that corresponds to

XX nucleotides 205-720 of the EMBL sequence J05581 with two

XX stop codons"

XX WO200025827-A2.

XX 11-MAY-2000.

XX 18-OCT-1999; 99WO-EP007874.

XX 30-OCT-1998; 98IT-MI002330.

XX (MENA) MENARINI RICERCHE SPA.

XX Parente D, Di Massimo AM, De Santis R;

XX WPI; 2000-365410/31.

XX P-PSDB; AAY71027.

XX Composition containing one or more DNA molecules encoding fragments of a

XX Mucin 1 (MUC-1) protein overexpressed in tumor cells, useful in anti-

XX tumor therapy.

XX Claim 18; Fig 8; 56pp; English.

XX The present sequence is a DNA encoding a fusion protein consisting of

XX human Mucin 1 (MUC-1) fragment fused to UBILacI sequence at the N-

XX terminus. The UBILacI sequence consists of ubiquitin from MCF7 cell line

XX and a portion of E. coli beta-galactosidase (LacI). MUC-1 is an antigenic

XX protein overexpressed in tumour cells. The present sequence is cloned

XX into a pMRS30 expression vector and used in pharmaceutical composition

XX e.g. vaccine for inducing an antigen-specific anti-tumour immune

XX response. Composition containing this DNA molecule is useful in anti-

XX tumour therapy of patients affected with tumours characterised by high

XX MUC-1 expression. (Updated on 15-SEP-2003 to standardise OS field)

XX SQ Sequence 891 BP; 203 A; 307 C; 225 G; 156 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 0.00103 Length: 891
Score: 109.00 Matches: 20
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 3 Gaps: 0

US-09-606-910E-2 (1-20) x AAD00391 (1-891)

Qy 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
Db 643 CACGGTGTACCTCGGCTCCGGATACAGCGCGGCCAGGTAGTACCGCCCTCTGCC 702

RESULT 8

ID AD287487
AC AD287487
XX
XX
XX 14-JUL-2005 (first entry)
XX
XX Human MUC1-EC protein-encoding gene SeqID14.
XX protein engineering; immunoglobulin; cytostatic; cancer; gene; ds.
XX Homo sapiens.
XX
XX WO2005042573-A1.
XX
XX 12-MAY-2005.
XX
XX 21-OCT-2004; 2004WO-US034680.
XX
XX 24-OCT-2003; 2003US-0514198P.
XX
XX 12-NOV-2003; 2003US-0519822P.
XX
XX (DAND) DANA FARBER CANCER INST INC.
XX
XX (ILEX-) ILEX PROD INC.
XX
XX Kharbanda S, Kufe DW;
XX
XX WPI; 2005-346855/35.
XX
XX P-PSDB; AD287486.

MUC1 ligand trap comprising a MUC1 extracellular polypeptide and an immunoglobulin FC polypeptide or an albumin polypeptide, useful for treating cancer.

PS Disclosure; SEQ ID NO 14; 82pp; English.

XX This invention relates to a novel MUC1 chimeric protein which comprises a first polypeptide sequence and a second polypeptide sequence, where the first polypeptide sequence is a MUC1 extracellular (MUC1-EC) polypeptide, and the second polypeptide sequence is a human immunoglobulin FC polypeptide or a human albumin polypeptide. The invention may be useful for the development of compounds with a cytostatic activity acting as MUC1 antagonists or CD227 antagonists. The MUC1 chimeric protein is useful in preparing a composition for treating cancer. The present sequence is that of a gene which was used during the development of the novel MUC1 chimeric protein of the invention.

XX Sequence 1125 BP; 240 A; 402 C; 246 G; 237 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 0.0013 Length: 1125
Score: 109.00 Matches: 20
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 14 Gaps: 0

US-09-606-910E-2 (1-20) x AD287487 (1-1125)

Qy 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
Db 406 CACGGTGTACCTCGGCTCCGGACACAGGCGCGGCCGCTCCACGCCCCCAGCC 465

RESULT 9

ID ADI57712 standard; cDNA; 1194 BP.
XX
XX ADI57712;
XX
XX 22-APR-2004 (first entry)
XX
XX Human breast specific nucleic acid (BSNA) #83.
XX
XX Human; breast specific nucleic acid; BSNA; gene; ss; metastasis;
XX breast cancer; cytostatic.
XX
XX Homo sapiens.
XX
XX WO2003106648-A2.
XX
XX 24-DEC-2003.
XX
XX 16-JUN-2003; 2003WO-US018934.
XX
XX 14-JUN-2002; 2002US-0389327P.
XX
XX (DIAD-) DIADEXUS INC.
XX
XX Salceda S, Macina RA, Turner LR, Sun Y, Liu C;
XX
XX WPI; 2004-082185/08.
XX
XX P-PSDB; ADI57782.

PT Novel isolated polypeptide comprising breast specific protein sequences, useful for diagnosing or monitoring presence and metastases of breast cancer in patient.

PS Claim 1; SEQ ID NO 83; 370pp; English.

XX The invention relates to human breast specific nucleic acids (BSNA) and the breast specific proteins (BSP) they encode. The nucleic acids are useful for determining the presence of a BSNA in a sample which involves contacting the sample with a BSNA under conditions in which the BSNA will selectively hybridise to a BSNA in the sample, and detecting the presence of a BSP in a sample which involves contacting the sample with suitable reagent under conditions in which the reagent will selectively interact with the BSP, and detecting the interaction of the reagent with a BSP in the sample. The nucleic acids and proteins are useful for diagnosing or monitoring the presence and metastases of breast cancer in a patient, which involves determining an amount of nucleic acid or protein in the sample of the patient to the amount of a breast specific marker in a normal control, where a difference in the determined amount in the sample compared to the amount in the control is associated with the presence of breast cancer. The sequences are useful for treating a patient with breast cancer, involving administering a composition consisting of a BSNA or a BSP to a patient, where the administration induces an immune response against the breast cancer cell expressing the BSNA or BSP. This sequence represents a human BSNA of the invention.

XX Sequence 1194 BP; 243 A; 438 C; 275 G; 238 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 0.00139 Length: 1194
Score: 109.00 Matches: 20
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 12 Gaps: 0

US-09-606-910E-2 (1-20) x ADI57712 (1-1194)

QY	1	HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20	
Db	617	CACGGTGTCACCTCGGCCCGGACACACGAGCGCGGCTCCACGGCCCCCAGCC 676	
RESULT 10			
AD	AAD00388	standard; DNA; 1371 BP.	
XX	AC	AAD00388;	
XX	DT	29-AUG-2000 (first entry)	
XX	DE	Human Mucin 1 (MUC-1) protein fragment encoding DNA #5.	
XX	KW	Human; Mucin 1; tumour; pMRS30 expression vector; anti-tumour;	
XX	KW	therapy; immune response; cytostatic; vaccine; ds.	
XX	OS	Homo sapiens.	
XX	FH	Key	Location/Qualifiers
XX	FT	CDS	1..1371
XX	FT	/*tag= a	
XX	FT	/product= "MUC-1 protein fragment"	
XX	PN	WO200025827-A2.	
XX	PD	11-MAY-2000.	
XX	XX	18-OCT-1999; 99WO-EP007874.	
XX	PF	30-OCT-1998; 98IT-MI002330.	
XX	PR	(MENA) MENARINI RICERCHE SPA.	
XX	PA	Parente D, Di Massimo AM, De Santis R;	
XX	PI	WPI; 2000-365410/31.	
XX	DR	P-PSDB; AAY71024.	
XX	CC	Composition containing one or more DNA molecules encoding fragments of a	
XX	CC	Mucin 1 (MUC-1) protein overexpressed in tumor cells, useful in anti-	
XX	CC	tumor therapy.	
XX	PS	Claim 16; Fig 5; 56pp; English.	
XX	CC	The present sequence is a DNA encoding a fragment of human Mucin 1 (MUC-	
XX	CC	1) antigenic protein which is overexpressed in tumour cells. The sequence	
XX	CC	was obtained by PCR from plasmids pMRS166, pMRS167, pMRS168 and pMRS169	
XX	CC	which contain MUC-1 DNA from BT20 tumour cells. It corresponds to	
XX	CC	nucleotides 136-1497 of the EMBL sequence J05581 with a start codon and	
XX	CC	two stop codons. The present sequence is cloned into a pMRS30 expression	
XX	CC	vector and used in pharmaceutical composition e.g. vaccine for inducing	
XX	CC	an antigen-specific anti-tumour immune response. Composition containing	
XX	CC	this DNA molecule is useful in anti-tumour therapy of patients affected	
XX	CC	with tumours characterised by high MUC-1 expression	
XX	SQ	Sequence 1371 BP; 300 A; 450 C; 321 G; 300 T; 0 U; 0 Other;	
Alignment Scores:			
Pred. No.:	0.00159	Length:	1371
Score:	109.00	Matches:	20
Percent Similarity:	100.0%	Conservative:	0
Best Local Similarity:	100.0%	Mismatches:	0
Query Match:	100.0%	Indels:	0
DB:	3	Gaps:	0
US-09-606-910E-2 (1-20) x AAD00388 (1-1371)			
QY	1	HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20	
Db	346	CACGGGTGTACCTCGGCTCCGGATACAGCGCGGCCAGGTAGTACCGCCCTCTGTC 405	
RESULT 11			
AD	ADI57693	standard; cDNA; 1378 BP.	
XX	AC	ADI57693;	
XX	DT	22-APR-2004 (first entry)	
XX	DE	Human breast specific nucleic acid (BSNA) #64.	
XX	KW	Human; breast specific nucleic acid; BSNA; gene; ss; metastasis;	
XX	KW	breast cancer; cytostatic.	
XX	OS	Homo sapiens.	
XX	PN	WO2003106648-A2.	
XX	XX	24-DEC-2003.	
XX	PF	16-JUN-2003; 2003WO-US018934.	
XX	PR	14-JUN-2002; 2002US-0389327P.	
XX	PA	(DIAD-) DIADEXUS INC.	
XX	PI	Salceda S, Macina RA, Turner LR, Sun Y, Liu C;	
XX	DR	WPI; 2004-082185/08.	
XX	DR	P-PSDB; ADI57765.	
XX	CC	Novel isolated polypeptide comprising breast specific protein sequences,	
XX	CC	useful for diagnosing or monitoring presence and metastases of breast	
XX	CC	cancer in patient.	
XX	PS	Claim 1; SEQ ID NO 64; 370pp; English.	
XX	CC	The invention relates to human breast specific nucleic acids (BSNA) and	
XX	CC	the breast specific proteins (BSP) they encode. The nucleic acids are	
XX	CC	useful for determining the presence of a BSNA in a sample which involves	
XX	CC	contacting the sample with a BSNA under conditions in which the BSNA will	
XX	CC	selectively hybridise to a BSNA in the sample, and detecting the	
XX	CC	hybridisation. The nucleic acids are useful for determining the presence	
XX	CC	of a BSP in a sample which involves contacting the sample with suitable	
XX	CC	reagent under conditions in which the reagent will selectively interact	
XX	CC	with the BSP, and detecting the interaction of the reagent with a BSP in	
XX	CC	the sample. The nucleic acids and proteins are useful for diagnosing or	
XX	CC	monitoring the presence and metastases of breast cancer in a patient,	
XX	CC	which involves determining an amount of nucleic acid or protein and	
XX	CC	comparing the determined amount of nucleic acid or protein in the sample	
XX	CC	of the patient to the amount of a breast specific marker in a normal	
XX	CC	control, where a difference in the determined amount in the sample	
XX	CC	compared to the amount in the control is associated with the presence of	
XX	CC	breast cancer. The sequences are useful for treating a patient with	
XX	CC	breast cancer, involving administering a composition consisting of a BSNA	
XX	CC	or a BSP to a patient, where the administration induces an immune	
XX	CC	response against the breast cancer cell expressing the BSNA or BSP. This	
XX	CC	sequence represents a human BSNA of the invention.	
XX	SQ	Sequence 1378 BP; 295 A; 489 C; 316 G; 278 T; 0 U; 0 Other;	
Alignment Scores:			
Pred. No.:	0.0016	Length:	1378
Score:	109.00	Matches:	20
Percent Similarity:	100.0%	Conservative:	0
Best Local Similarity:	100.0%	Mismatches:	0
Query Match:	100.0%	Indels:	0
DB:	12	Gaps:	0
US-09-606-910E-2 (1-20) x ADI57693 (1-1378)			
QY	1	HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20	
Db	617	CACGGGTGTACCTCGGCTCCGGACACACGAGCGCGGCTCCACGGCCCCCAGCC 676	


```
Db      406 CACGGTGTACCTCGGCCCGGACACACGAGCGCGCGCTCCACCGCCCCCGAGCC 465
RESULT 14
ADO23125
ID      ADO23125 standard; RNA; 1428 BP.
XX      AC
XX      ADO23125;
XX      DT
XX      12-AUG-2004 (first entry)
XX      DE Human MUC1 mucin glycoprotein RNA (coding sequence) SeqID 20.
XX      KW human; MUC1; mucin glycoprotein; cancer; chemotherapeutic; MUC1/ECD; ss.
XX      OS Homo sapiens.
XX      PN WO200404160-A2.
XX      PD 27-MAY-2004.
XX      PF 12-NOV-2003; 2003WO-US035848.
XX      PR 13-NOV-2002; 2002US-00293391.
XX      PR 29-MAY-2003; 2003US-00447839.
XX      PA (DAND ) DANA FARMER CANCER INST INC.
XX      PA (ILEX-) ILEX PROD INC.
XX      PI Kufe DW, Kharbanda S, Weitman SD;
XX      WI WI; 2004-420304/39.
XX      DR Double-stranded RNA complex useful for inhibiting proliferation of cancer
XX      PT cell expressing MUC1 mucin glycoprotein, comprises first and second RNA
XX      PT sequences.
XX      PS Claim 2; SEQ ID NO 20; 112pp; English.
XX      CC This invention relates to novel modulators of the human MUC1 mucin
XX      CC glycoprotein for use in cancer therapeutics, where MUC1 is a protein that
XX      CC acts to inhibit the apoptotic response to genotoxic stress caused by
XX      CC chemotherapeutic agents. In particular, it refers to modulators of the
XX      CC MUC1 extracellular domain (MUC1/ECD). The method refers to using double-
XX      CC stranded RNA complexes as MUC1 interference RNA compositions such that
XX      CC MUC1 expression is inhibited, which in turn inhibits cancer cell
XX      CC proliferation. The present invention describes screening assays to
XX      CC identify compounds that inhibit the binding of various MUC1 ligands such
XX      CC as neuregulin 2, as well as agonists, antagonists and antibodies thereof.
XX      CC Furthermore, it provides MUC1 antisense, small interfering RNA (siRNA)
XX      CC and small molecules in combination with chemotherapeutic agents that are
XX      CC useful in the field of cancer therapy. This polynucleotide sequence is
XX      CC the human MUC1 RNA of the invention.
XX      SQ Sequence 1428 BP; 296 A; 496 C; 327 G; 0 T; 309 U; 0 Other;

Alignment Scores:
Pred. No.: 0.00166 Length: 1428
Score: 109.00 Matches: 20
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 12 Gaps: 0

US-09-606-910E-2 (1-20) x ADO23125 (1-1428)

QY      1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
Db      406 CACGGUGUACACCGCGCGCGGACACACGAGCGCGCGCGGTCUCCACCGCCCCCGAGCC 465
RESULT 15
ADF32627
ID      ADF32627 standard; DNA; 1457 BP.
```

```
XX      ADF32627;
AC      26-FEB-2004 (first entry)
XX      DT
XX      DE Plasmid JNW358 MUC-1 nucleotide sequence.
XX      KW MUC-1 antigen; immune response; MUC-1; variable number of tandem repeat;
XX      KW VNTR; repeat unit; tumour; metastasis; cytostatic; vaccine; gene therapy;
XX      KW gene; ds.
XX      OS Synthetic.
XX      PN WO2003100060-A2.
XX      PD 04-DEC-2003.
XX      PF 23-MAY-2003; 2003WO-EP005594.
XX      PR 24-MAY-2002; 2002GB-00012046.
XX      PA (GLAX ) GLAXO GROUP LTD.
XX      PI Burden N, Ellis JH, Hamblin PA;
XX      WI WI; 2004-042811/04.
XX      DR New nucleic acid molecule encoding a MUC-1 antigen, useful for preparing
XX      PT a composition for treating or preventing tumors or metastases.
XX      PT Example; Fig 3; 66pp; English.
XX      PS The present invention describes a nucleic acid molecule which encodes a
XX      CC MUC-1 antigen. The nucleic acid is capable of raising an immune response
XX      CC in vivo, has reduced susceptibility to recombination than full-length MUC
XX      CC -1 and comprises between 1 and 15 variable number of tandem repeats
XX      CC (VNTR) perfect repeat units. Also described: (1) a plasmid comprising the
XX      CC DNA molecule; (2) a protein encoded by the nucleic acid; (3) a
XX      CC pharmaceutical composition comprising the nucleic acid, plasmid or
XX      CC protein and an excipient, diluent or carrier; and (4) a method of
XX      CC treating or preventing tumors or metastases. A MUC1 antigen has
XX      CC cytostatic activity, and can be used in vaccines, and in gene therapy.
XX      CC The nucleic acid is useful for preparing a composition for treating or
XX      CC preventing tumors or metastases. The present sequence is used in the
XX      CC exemplification of the present invention.
XX      SQ Sequence 1457 BP; 302 A; 506 C; 334 G; 315 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 0.0017 Length: 1457
Score: 109.00 Matches: 20
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 12 Gaps: 0

US-09-606-910E-2 (1-20) x ADF32627 (1-1457)

QY      1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
Db      424 CACGGTGTACCTCGGCCCGGACACACGAGCGCGCGCGGTCCACCGCCCCCGAGCC 483

Search completed: April 14, 2006, 19:06:33
Job time : 368.78 secs
```

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 14, 2006, 18:49:30 ; Search time 3146.83 Seconds
(without alignments)
297.360 Million cell updates/sec

Title: US-09-606-910E-2

Perfect score: 109
Sequence: 1 HGVTSAPTRPAGSTAPPA 20

Scoring table:
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlp
-Q=/abes/ABSWEB spool/US09606910/runat 14042006 091621 1556/app query.fasta_1
-DB=EST -QFMT=fasta -SUPPIX=p2n.rst -MINMATCH=0.1 -LOOFCU=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MAIRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THRM SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abss05p
-USER=US09606910 @CCN 1 1 10172 @runat 14042006 091621 1556 -NCPU=6 -ICPU=3
-NO MMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_hic:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	109	100.0	330	1	AI925867 wo20d04.x
2	109	100.0	604	3	BM791359 K-EST0071
3	109	100.0	754	8	DR422781 nav17d09
4	109	100.0	877	5	BU542454 AGENCOURT
5	109	100.0	959	6	CA489836 AGENCOURT
6	109	100.0	1113	5	BU148487 AGENCOURT
7	109	100.0	1130	5	BU542996

8	109	100.0	1234	5	BQ935898
9	109	100.0	1262	5	BQ935496
10	109	100.0	1343	5	BQ920055
11	109	100.0	1349	5	BU152566
12	109	100.0	1420	5	BU542790
13	109	100.0	1531	5	BU543309
14	101	92.7	1268	5	BQ943554
15	100	91.7	1334	5	BQ943809
16	90	82.6	472	3	BM759495
17	90	82.6	475	6	CB120860
18	90	82.6	669	6	CB122585
19	90	82.6	981	2	BG774910
20	90	82.6	1536	5	BQ923149
21	86	78.9	166	8	T27692
22	81	74.3	619	2	B1260921
23	79	72.5	1678	2	BG775565
24	76	69.7	690	1	AL543598
25	76	69.7	1241	4	CR596859
26	67	61.5	1187	3	BQ219669
27	67	61.5	1366	3	BM046583
c 28	65	59.6	172	2	B1015399
c 29	65	59.6	940	2	BE901435
c 30	64.5	59.2	333	10	CL974028
31	64	58.7	154	2	BF463559
32	64	58.7	344	1	AW494383
33	63.5	58.3	533	3	B1795528
c 34	63	57.8	472	10	CL804119
35	63	57.8	548	2	BE706360
c 36	63	57.8	708	3	BM582005
37	63	57.8	722	7	CR993345
38	63	57.8	786	2	BG718241
39	62	56.9	483	7	CK061420
40	62	56.9	500	1	AV591193
c 41	62	56.9	549	2	BF929925
42	62	56.9	622	7	CV014984
c 43	62	56.9	656	1	AV609791
c 44	62	56.9	831	10	CZ356841
c 45	62	56.9	956	10	CNS02534

ALIGNMENTS

RESULT 1
AI925867/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AI925867 330 bp mRNA linear EST 08-MAR-2000
wo20d04.x1 NCI CGAP Panel Homo sapiens cDNA clone IMAGE:2455879 3'
similar to contains element MSRI MSRI repetitive element ;, mRNA
sequence.
AI925867 GI:5661831
EST.
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
1 (bases 1 to 330)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
CONTACT: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Life Technologies catalog #: 11548-013
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 578 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 305.
Location/Qualifiers
1. .330

cdna was synthesized from oligo dT-selected mRNA by priming with dT-tailed vector. The dT-tailed vector was adjusted to have about 60nt. The cdna vector was circularized with E. coli DNA ligase after digestion of EcoRI which site is also included in vector. An RNA strand converted to a DNA strand by Okayama-Berg method. The obtained cdna vectors were used for transfection of competent cells E. coli Top10p by electroporation method. The cdna libraries constructed by this method are full-length enriched cdna library."

ORIGIN

Alignment Scores:
 Pred. No.: 0.00551 Length: 604
 Score: 109.00 Matches: 20
 Percent Similarity: 100.0% Conservative: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 100.0% Indels: 0
 DB: 3 Gaps: 0

US-09-606-910E-2 (1-20) x BM791359 (1-604)

Qy 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
 Db 127 CACGGTGTCACTCGGCCCGGACACACAGCGCGCGCGGCTCCACGCCCCCAGCC 186

RESULT 3

DR422781

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Location/Qualifiers

1. .754

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="nav17d09"

/tissue_type="Pterygium"

/dev_stage="Adult"

/lab_host="EMDH10B"

/clone_lib="Human pterygium. Unnormalized (nav)"

/note="Organ: Eye; Vector: pCMVSPORT6; RNA was extracted from 9 pooled human pterygia. A directionally cloned cdna library in the pCMVSPORT6 vector (Invitrogen) was constructed at Bioserve Biotechnology (laurel MD) essentially following the protocols of the SuperScript Plasmid System, full details of which are contained in the manufacturer's instruction manual (http://www.lifetech.com/). First strand synthesis was carried out using a Not I primer-adaptor [5'-pGACTAGTTCTAGATCGGCGCGGCC(T)15-3']. cdna was

Seq primer: Universal M13 Reverse.

Plate: 17 row: d column: 09

Fax: 301 496 0078

Email: graeme@helix.nih.gov

Tel: 301 402 3452

6/331, NIH, Bethesda, MD 20892-2740, USA

National Eye Institute

Section on Molecular Structure and Function

Contact: Wistow G

Unpublished (2005)

NEIBank analysis of Human pterygium

Reid, T., Dushku, N. and Carper, D.

Wistow, G., Jaworski, C., Aryankalayil-John, M., Rowsey, J.J., Cox, C.,

1 (bases 1 to 754)

Hominidae; Homo.

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:2455879"

/tissue_type="adenocarcinoma"

/lab_host="DH10B"

/clone_lib="NCI_CGAP_Pan1"

/note="Organ: pancreas; Vector: pCMV-SPORT6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.

Average insert size 1.72 kb. Life Technologies catalog #:

11548-013"

Alignment Scores:
 Pred. No.: 0.00307 Length: 330
 Score: 109.00 Matches: 20
 Percent Similarity: 100.0% Conservative: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 100.0% Indels: 0
 DB: 1 Gaps: 0

US-09-606-910E-2 (1-20) x A1925867 (1-330)

Qy 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
 Db 295 CACGGTGTCACTCGGCCCGGACACACAGCGCGCGGCTCCACGCCCCCAGCC 236

RESULT 2

BM791359

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Kim YS

Genome Research Center

Korea Research Institute of Bioscience & Biotechnology

52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea

Tel: +82-42-860-4470

Fax: +82-42-860-4409

Email: yongsung@mail.kribb.re.kr

Plate: 14 row: A column: 06

High quality sequence stop: 604.

Location/Qualifiers

1. 604

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="S21SNU520-14-A06"

/sex="F"

/tissue_type="Stomach"

/cell_type="Floating aggregates"

/cell_line="SNU-520"

/lab_host="Top10F"

/clone_lib="S21SNU520"

/note="Organ: Stomach; Vector: pTZ18RP1; Site 1: EcoRI; Site 2: NotI. The poly (A) + RNA was dephosphorylated with bacterial alkaline phosphatase (BAP) and then decapped with tabacco acid pyrophosphatase (TAP). The decapped intact mRNA was ligated with DNA-RNA linker including EcoR

I site by treatment of T4 RNA ligase and the first strand

cloned in Not I/Sal I sites. EST analysis was performed at the NIH Intramural Sequencing Center (NISC). Analyzed data available through <http://neibank.nei.nih.gov>."

ORIGIN

Alignment Scores:
Pred. No.: 0.00682 Length: 754
Score: 109.00 Matches: 20
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 8 Gaps: 0

US-09-606-910E-2 (1-20) x DR422781 (1-754)

QY 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
|||||
DB 57 CACGGTGTCACTCGGCCCGGACACACAGCGCGCGGGCTCCACCGACCCCGCAGCC 116

RESULT 4
LOCUS BU542454 877 bp mRNA linear EST 13-SEP-2002
DEFINITION AGENCOURT_10322173 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6574322
5', mRNA sequence.

ACCESSION BU542454
VERSION BU542454.1 GI:22852937
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.

REFERENCE 1 (bases 1 to 877)
AUTHORS NIH-MGC <http://mgs.nci.nih.gov/>
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DTP

CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: LLCM2769 row: h column: 02
High quality sequence stop: 760.

FEATURES

source

1. .877
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6574322"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC 40"
/note="Organ: prostate; Vector: pOTB7; Site 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC library."

ORIGIN

Alignment Scores:
Pred. No.: 0.00789 Length: 877
Score: 109.00 Matches: 20
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 5 Gaps: 0

US-09-606-910E-2 (1-20) x BU542454 (1-877)

QY 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
|||||
DB 56 CACGGTGTCACTCGGCCCGGACACACAGCGCGCGGGCTCCACCGACCCCGCAGCC 115

RESULT 5
LOCUS CA489836 959 bp mRNA linear EST 14-NOV-2002
DEFINITION AGENCOURT_10810668 MAPCL Homo sapiens cDNA clone IMAGE:6722324 5',
mRNA sequence.

ACCESSION CA489836
VERSION CA489836.1 GI:24952627
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.

REFERENCE 1 (bases 1 to 959)
AUTHORS NIH-MGC <http://mgs.nci.nih.gov/>
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Kristi A. Eglund, Ira Pastan
CDNA Library Preparation: Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: LLAM14284 row: n column: 20
High quality sequence start: 31
High quality sequence stop: 446.

FEATURES

source

1. .959
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6722324"
/cell_line="ZR-75-1, MCF7, SK-BR-3, MDA-MB-231,
HRT-EMEI, LNCap"
/lab_host="EMDH10B"
/clone_lib="MAPCL"
/note="Vector: pCMV-SPORT6; Site 1: EcoRV; Site 2: Not I;
Subtracted with brain, liver, lung, kidney and muscle.
Directionally cloned. Priming method: oligo-dT. Average
insert size: 1800 bp. Library amplification: 26,000 fold.
Kristi A. Eglund, James J. Vincent, Robert Strausberg,
Bungkook Lee & Ira Pastan: Discovery of new breast
cancer genes encoding membrane and secreted proteins.
Manuscript submitted."

ORIGIN

Alignment Scores:
Pred. No.: 0.0086 Length: 959
Score: 109.00 Matches: 20
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 6 Gaps: 0

US-09-606-910E-2 (1-20) x CA489836 (1-959)

QY 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
|||||
DB 430 CACGGTGTCACTCGGCCCGGACACACAGCGCGCGGGCTCCACCGACCCCGCAGCC 489

RESULT 6

BU148487 1113 bp mRNA linear EST 03-SEP-2002
LOCUS

```

DEFINITION  AGENCOURT_8670479 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6380642
5', mRNA sequence.
ACCESSION   BUI48487
VERSION     BUI48487.1 GI:22662019
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homnidae; Homo.
REFERENCE   1 (bases 1 to 1113)
AUTHORS    NIH-MGC http://mgc.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL    Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: DCTD/DRP
            cDNA Library Preparation: Rubin Laboratory
            DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLC2569 row: j column: 03
            High quality sequence stop: 235.
FEATURES    Location/Qualifiers
             1..1113
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:6380642"
                /tissue_type="carcinoma, cell line"
                /lab_host="DH10B (phage-resistant)"
                /clone_lib="NIH_MGC_40"
                /note="Organ: prostate; Vector: pOTB7; Site 1: XhoI;
                Site 2: EcoRI; cDNA made by oligo-dT priming.
                Directionally cloned into EcoRI/XhoI sites using the
                following 5' adaptor: GGCACGAG(G). Library constructed by
                Ling Hong in the laboratory of Gerald M. Rubin (University
                of California, Berkeley) using ZAP-cDNA synthesis kit
                (Stratagene) and Superscript II RT (Life Technologies).
                Note: this is a NIH_MGC Library."
ORIGIN
Alignment Scores:
Pred. No.:      0.00993      Length:      1113
Score:          109.00      Matches:    20
Percent Similarity: 100.0%      Conservative: 0
Best Local Similarity: 100.0%      Mismatches: 0
Query Match:    100.0%      Indels:     0
DB:              5          Gaps:         0

US-09-606-910E-2 (1-20) x BUI48487 (1-1113)

QY      1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProAla 20
Db      51 CACGGTGTCACTCTGCCCGGACACACGCGCGCGGCTCCACCGCCCCCAGCC 110

RESULT 7
BU542996
LOCUS     BU542996
DEFINITION  BU542996 1130 bp mRNA linear EST 13-SEP-2002
5', mRNA sequence.
ACCESSION   BU542996
VERSION     BU542996.1 GI:22853479
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homnidae; Homo.
REFERENCE   1 (bases 1 to 1130)
AUTHORS    NIH-MGC http://mgc.nci.nih.gov/.

```

```

TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL    Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: DCTD/DRP
            cDNA Library Preparation: Rubin Laboratory
            DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLC2771 row: c column: 11
            High quality sequence start: 27
            High quality sequence stop: 246.
FEATURES    Location/Qualifiers
             1..1130
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:6574979"
                /tissue_type="carcinoma, cell line"
                /lab_host="DH10B (phage-resistant)"
                /clone_lib="NIH_MGC_40"
                /note="Organ: prostate; Vector: pOTB7; Site 1: XhoI;
                Site 2: EcoRI; cDNA made by oligo-dT priming.
                Directionally cloned into EcoRI/XhoI sites using the
                following 5' adaptor: GGCACGAG(G). Library constructed by
                Ling Hong in the laboratory of Gerald M. Rubin (University
                of California, Berkeley) using ZAP-cDNA synthesis kit
                (Stratagene) and Superscript II RT (Life Technologies).
                Note: this is a NIH_MGC Library."
ORIGIN
Alignment Scores:
Pred. No.:      0.0101      Length:      1130
Score:          109.00      Matches:    20
Percent Similarity: 100.0%      Conservative: 0
Best Local Similarity: 100.0%      Mismatches: 0
Query Match:    100.0%      Indels:     0
DB:              5          Gaps:         0

US-09-606-910E-2 (1-20) x BU542996 (1-1130)

QY      1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProAla 20
Db      90 CACGGTGTCACTCTGCCCGGACACACGCGCGCGGCTCCACCGCCCCCAGCC 149

RESULT 8
BO936898
LOCUS     BO936898
DEFINITION  BO936898 1234 bp mRNA linear EST 21-AUG-2002
5', mRNA sequence.
ACCESSION   BO936898
VERSION     BO936898.1 GI:22352281
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homnidae; Homo.
REFERENCE   1 (bases 1 to 1234)
AUTHORS    NIH-MGC http://mgc.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL    Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: DCTD/DRP
            cDNA Library Preparation: Rubin Laboratory
            DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov

```

Plate: LLCM2577 row: 0 column: 24
High quality sequence stop: 245.

FEATURES

Location/Qualifiers
1. .1234
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6383855"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_40"
/note="Organ: prostate; Vector: pOTB7; Site: 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:
Pred. No.: 0.011 Length: 1234
Score: 109.00 Matches: 20
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 5 Gaps: 0

US-09-606-910E-2 (1-20) x BQ936898 (1-1234)

Qy 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
|||||
Db 51 CACGGTGTACCTCTGCCCGGACACGAGCGCGCGGCTCCACCGCCCCCAGCC 110

RESULT 9

BQ935496
LOCUS BQ935496 1262 bp mRNA linear EST 21-AUG-2002
DEFINITION AGENCOURT 8776475 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6382088
5' mRNA sequence.

ACCESSION BQ935496

VERSION BQ935496.1 GI:22350879

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.

1 (bases 1 to 1262)

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgaabs-remail.nih.gov

Tissue Procurement: DCTD/DTF

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLCM2573 row: f column: 09

High quality sequence start: 46

High quality sequence stop: 157.

Location/Qualifiers

1. .1262

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:6382088"

/tissue_type="carcinoma, cell line"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NIH_MGC_40"

FEATURES

source

/note="Organ: prostate; Vector: pOTB7; Site: 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:
Pred. No.: 0.0112 Length: 1262
Score: 109.00 Matches: 20
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 5 Gaps: 0

US-09-606-910E-2 (1-20) x BQ935496 (1-1262)

Qy 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
|||||
Db 77 CACGGTGTACCTCTGCCCGGACACGAGCGCGCGGCTCCACCGCCCCCAGCC 136

RESULT 10

BQ920055
LOCUS BQ920055 1343 bp mRNA linear EST 20-AUG-2002
DEFINITION AGENCOURT 10031674 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6481609
5' mRNA sequence.

ACCESSION BQ920055

VERSION BQ920055.1 GI:22334753

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.

1 (bases 1 to 1343)

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgaabs-remail.nih.gov

Tissue Procurement: DCTD/DTF

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLCM2663 row: e column: 02

High quality sequence start: 56

High quality sequence stop: 237.

Location/Qualifiers

1. .1343

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:6481609"

/tissue_type="carcinoma, cell line"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NIH_MGC_40"

/note="Organ: prostate; Vector: pOTB7; Site: 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming.

Directionally cloned into EcoRI/XhoI sites using the

following 5' adaptor: GGCACGAG(G). Library constructed by

Ling Hong in the laboratory of Gerald M. Rubin (University

of California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies).

Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:

```

Pred. No.: 0.0119 Length: 1343
Score: 109.00 Matches: 20
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 5 Gaps: 0

US-09-606-910E-2 (1-20) x BQ920055 (1-1343)

Qy 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
Db 93 CACGGTGTACCTCTGCGCCGGACACCGAGCGCGCGGGCTCCACGCCCGCCCGAGCC 152

RESULT 11
BU152566 1349 bp mRNA linear EST 03-SEP-2002
LOCUS AGENCOURT_8670683 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6380559
DEFINITION 5', mRNA sequence.
ACCESSION BU152566
VERSION BU152566.1 GI:22666098
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 1349)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LNCM2569 row: f column: 16
High quality sequence stop: 291.
Location/Qualifiers
1. .1349
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6380559"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 40"
/notes="Organ: prostate; Vector: pOTB7; Site 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
Note: this is a NIH_MGC Library."

FEATURES
source
1. .1349
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6380559"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 40"
/notes="Organ: prostate; Vector: pOTB7; Site 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
Note: this is a NIH_MGC Library."

ORIGIN
Alignment Scores:
Pred. No.: 0.012 Length: 1349
Score: 109.00 Matches: 20
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 5 Gaps: 0

US-09-606-910E-2 (1-20) x BU152566 (1-1349)

Qy 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
Db 40 CACGGTGTACCTCTGCGCCGGACACCGAGCGCGGGCTCCACGCCCGCCCGAGCC 99

RESULT 13
BU543309 1531 bp mRNA linear EST 13-SEP-2002
LOCUS AGENCOURT_10327072 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6575349
DEFINITION 5', mRNA sequence.
ACCESSION BU543309
VERSION BU543309.1 GI:22853792
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Hominidae; Homo.
1 (bases 1 to 1531)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LNCM2770 row: h column: 21
High quality sequence stop: 288.
Location/Qualifiers
1. .1420
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6574725"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 40"
/notes="Organ: prostate; Vector: pOTB7; Site 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
Note: this is a NIH_MGC Library."

ORIGIN
Alignment Scores:
Pred. No.: 0.0126 Length: 1420
Score: 109.00 Matches: 20
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 5 Gaps: 0

US-09-606-910E-2 (1-20) x BU542790 (1-1420)

Qy 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
Db 40 CACGGTGTACCTCTGCGCCGGACACCGAGCGCGGGCTCCACGCCCGCCCGAGCC 99

```


Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 1531)
AUTHORS NIH-MGC <http://mgs.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTF
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LNCM2772 row: b column: 21
High quality sequence stop: 166.
Location/Qualifiers
1..1531
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6575349"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_40"
/note="Organ: prostate; Vector: pOTB7; Site 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

Alignment Scores:
Pred. No.: 0.0135 Length: 1531
Score: 109.00 Matches: 20
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 5 Gaps: 0

US-09-606-910E-2 (1-20) x BU543309 (1-1531)

QY 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProAla 20
DB 51 CACGGTGTCACTCTGCCCGGACACAGCGCGCGGCTCCACCGCCCCCAGCC 110

RESULT 14
BQ943554 1268 bp mRNA linear EST 21-AUG-2002
LOCUS AGENCOURT_8777815 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6384308
DEFINITION 5', mRNA sequence.
ACCESSION BQ943554
VERSION BQ943554.1 GI:22359032
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 1268)
AUTHORS NIH-MGC <http://mgs.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTF
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LNCM2579 row: b column: 21
High quality sequence stop: 177.
Location/Qualifiers
1..1268
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6384308"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_40"
/note="Organ: prostate; Vector: pOTB7; Site 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

Alignment Scores:
Pred. No.: 0.0965 Length: 1268
Score: 101.00 Matches: 18
Percent Similarity: 94.7% Conservative: 0
Best Local Similarity: 94.7% Mismatches: 1
Query Match: 92.7% Indels: 0
DB: 5 Gaps: 0

US-09-606-910E-2 (1-20) x BQ943554 (1-1268)

QY 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 19
DB 111 CACGGTGTCACTCTGCCCGGACACAGCGCGCGGCTCCACCGCCCCCA 167

RESULT 15
BQ943809 1334 bp mRNA linear EST 21-AUG-2002
LOCUS AGENCOURT_8775676 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6383286
DEFINITION 5', mRNA sequence.
ACCESSION BQ943809
VERSION BQ943809.1 GI:22359287
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 1334)
AUTHORS NIH-MGC <http://mgs.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTF
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LNCM2576 row: h column: 07
High quality sequence stop: 114.
Location/Qualifiers
1..1334
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6383286"
/tissue_type="carcinoma, cell line"

/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_40"
/note="Organ: prostate; Vector: pOTB7; Site 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GCCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:
Pred. No.: 0.133 Length: 1334
Score: 100.00 Matches: 18
Percent Similarity: 94.7% Conservative: 0
Best Local Similarity: 94.7% Mismatches: 1
Query Match: 91.7% Indels: 0
DB: 5 Gaps: 0

US-09-606-910E-2 (1-20) x BQ943809 (1-1334)

Qy 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 19
Db 53 CACGGTGTACCTCGGCCCCGACACACCGCGCGCGCGGCTCCACGCCCCCCCA 109

Search completed: April 14, 2006, 21:33:42
Job time : 3147.83 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 14, 2006, 18:54:01 ; Search time 116.585 Seconds

(without alignments)
304.937 Million cell updates/sec

Title: US-09-606-910E-2

Perfect score: 109

Sequence: 1 HGVTSPADTRPAGSTAPPA 20

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DEV=xlh

-MODE=frame+pn.model -DEV=xlh
-O=/abss/ABSSWEB.spool/US09606910/runat_14042006_091623_1610/app_query.fasta_1
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DLOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-HOST=abs808 -USER=US09606910 @CGN_1_1_237 @runat_14042006_091623_1610 -NCPU=6
-ICPU=3 -NO_MMAPP -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA:
1: /cgn2_6/prodata/1/ina/1 COMB.seq.*
2: /cgn2_6/prodata/1/ina/5 COMB.seq.*
3: /cgn2_6/prodata/1/ina/6A COMB.seq.*
4: /cgn2_6/prodata/1/ina/6B COMB.seq.*
5: /cgn2_6/prodata/1/ina/H COMB.seq.*
6: /cgn2_6/prodata/1/ina/PCTUS COMB.seq.*
7: /cgn2_6/prodata/1/ina/PP COMB.seq.*
8: /cgn2_6/prodata/1/ina/RE COMB.seq.*
9: /cgn2_6/prodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	109	100.0	1721	3	US-10-029-517-3
2	109	100.0	1804	3	Sequence 3, Appli
3	109	100.0	8186	3	Sequence 17, Appli
4	101	92.7	519	3	Sequence 19, Appli
5	101	92.7	534	3	Sequence 42, Appli
6	101	92.7	572	3	Sequence 46, Appli
7	94	86.2	518	3	Sequence 18, Appli
8	94	86.2	3343	3	Sequence 101, Appli
9	91	83.5	6192	2	Sequence 102, Appli
					Sequence 1, Appli

10	91	83.5	6192	3	US-09-083-116-1	Sequence 1, Appli
11	91	83.5	6192	3	US-09-134-916A-1	Sequence 1, Appli
12	91	83.5	6449	2	US-08-479-537A-4	Sequence 4, Appli
13	91	83.5	6449	3	US-09-083-116-4	Sequence 4, Appli
14	91	83.5	6449	3	US-09-134-916A-4	Sequence 4, Appli
15	90	82.6	981	3	US-10-029-517-16	Sequence 16, Appli
16	85	78.0	68	3	US-09-304-967-51	Sequence 51, Appli
17	84	77.1	83	3	US-09-304-967-75	Sequence 75, Appli
18	83	76.1	48	3	US-09-304-967-47	Sequence 47, Appli
19	83	76.1	48	3	US-09-304-967-67	Sequence 67, Appli
20	83	76.1	48	3	US-09-304-967-90	Sequence 90, Appli
21	83	76.1	68	3	US-09-304-967-49	Sequence 49, Appli
22	83	76.1	68	3	US-09-304-967-53	Sequence 53, Appli
23	83	76.1	68	3	US-09-304-967-55	Sequence 55, Appli
24	83	76.1	68	3	US-09-304-967-57	Sequence 57, Appli
25	83	76.1	78	3	US-09-304-967-92	Sequence 92, Appli
26	83	76.1	78	3	US-09-304-967-94	Sequence 94, Appli
27	83	76.1	78	3	US-09-304-967-96	Sequence 96, Appli
28	83	76.1	78	3	US-09-304-967-98	Sequence 98, Appli
29	83	76.1	78	3	US-09-304-967-100	Sequence 100, Appli
30	83	76.1	78	3	US-09-304-967-102	Sequence 102, Appli
31	83	76.1	83	3	US-09-304-967-69	Sequence 69, Appli
32	83	76.1	83	3	US-09-304-967-71	Sequence 71, Appli
33	83	76.1	83	3	US-09-304-967-73	Sequence 73, Appli
34	83	76.1	83	3	US-09-304-967-77	Sequence 77, Appli
35	83	76.1	83	3	US-09-304-967-79	Sequence 79, Appli
36	64.5	59.2	11220	3	US-09-105-537-32	Sequence 32, Appli
37	64.5	59.2	36778	3	US-09-105-537-5	Sequence 5, Appli
38	64.5	59.2	38506	3	US-09-320-878-19	Sequence 19, Appli
39	64.5	59.2	38506	3	US-09-141-908-1	Sequence 1, Appli
40	64.5	59.2	38506	3	US-09-657-440-19	Sequence 19, Appli
41	64.5	59.2	38506	3	US-09-793-708-19	Sequence 19, Appli
42	63.5	58.3	4403765	3	US-09-103-840A-2	Sequence 2, Appli
43	63.5	58.3	4411529	3	US-09-103-840A-1	Sequence 1, Appli
44	61	56.0	60	3	US-09-475-947A-246	Sequence 246, Appli
45	61	56.0	942	3	US-09-902-540-7357	Sequence 7357, Appli

ALIGNMENTS

RESULT 1

US-10-029-517-3
; Sequence 3, Application US/10029517
; Patent No. 6716627
; GENERAL INFORMATION:
; APPLICANT: Kenneth W. Dobie
; APPLICANT: Susan J. Myers
; TITLE OF INVENTION: ANTISENSE MODULATION OF MUCIN 1, TRANSMEMBRANE EXPRESSION
; FILE REFERENCE: RTS-0352
; CURRENT APPLICATION NUMBER: US/10/029,517
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 107
; SEQ ID NO 3
; LENGTH: 1721
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (58)....(1605)
US-10-029-517-3

Alignment Scores:
Pred. No.: 0.000276 Length: 1721
Score: 109.00 Matches: 20
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 3 Gaps: 0

US-09-606-910E-2 (1-20) x US-10-029-517-3 (1-1721)

QY 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProCala 20

|||||

```
Db      583 CACGGTGTCACTCGGCCCGGACACACAGGCGGGCGGCTCCACCGCCCCCGCAGCC 642
RESULT 2
US-10-029-517-17
; Sequence 17, Application US/10029517
; Patent No. 6716627
; GENERAL INFORMATION:
; APPLICANT: Kenneth W. Dobie
; APPLICANT: Susan J. Myers
; TITLE OF INVENTION: ANTISENSE MODULATION OF MUCIN 1, TRANSMEMBRANE EXPRESSION
; FILE REFERENCE: RTS-0352
; CURRENT APPLICATION NUMBER: US/10/029,517
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 107
; SEQ ID NO 17
; LENGTH: 1804
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (73)...(1500)
US-10-029-517-17
Alignment Scores:
Pred. No.:      0.00289      Length:      1804
Score:          109.00      Matches:      20
Percent Similarity: 100.0%      Conservative: 0
Best Local Similarity: 100.0%      Mismatches: 0
Query Match:    100.0%      Indels:      0
DB:             3          Gaps:          0

US-09-606-910E-2 (1-20) x US-10-029-517-17 (1-1804)
Qy      1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
Db      478 CACGGTGTCACTCGGCCCGGACACACAGGCGGGCGGCTCCACCGCCCCCGCAGCC 537
RESULT 3
US-10-029-517-19
; Sequence 19, Application US/10029517
; Patent No. 6716627
; GENERAL INFORMATION:
; APPLICANT: Kenneth W. Dobie
; APPLICANT: Susan J. Myers
; TITLE OF INVENTION: ANTISENSE MODULATION OF MUCIN 1, TRANSMEMBRANE EXPRESSION
; FILE REFERENCE: RTS-0352
; CURRENT APPLICATION NUMBER: US/10/029,517
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 107
; SEQ ID NO 19
; LENGTH: 8186
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 6899
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 7155
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 7184
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 7957
; OTHER INFORMATION: unknown
; NAME/KEY: intron
; LOCATION: (2997)...(3498)
; OTHER INFORMATION: intron 1
; NAME/KEY: intron:exon junction
; LOCATION: (3498)...(3499)
; OTHER INFORMATION: intron 1:exon 2
; NAME/KEY: exon

US-09-606-910E-2 (1-20) x US-10-029-517-19 (1-8186)
Qy      1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
Db      3846 CACGGTGTCACTCGGCCCGGACACACAGGCGGGCGGCTCCACCGCCCCCGCAGCC 3905
RESULT 4
US-09-646-028-42
; Sequence 42, Application US/09646028
; Patent No. 6562347
; GENERAL INFORMATION:
; APPLICANT: Kwak, Larry
; APPLICANT: Biragyn, Arya
; TITLE OF INVENTION: METHODS AND COMPOSITIONS OF
; FILE REFERENCE: 14014.0316/P
; CURRENT APPLICATION NUMBER: US/09/646,028
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 60/077,745
; PRIOR FILING DATE: 1998-03-12
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 42
; LENGTH: 519
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of artificial sequence:/note=synthetic construct
US-09-646-028-42
Alignment Scores:
Pred. No.:      0.00072      Length:      519
Score:          101.00      Matches:      19
Percent Similarity: 100.0%      Conservative: 0
Best Local Similarity: 100.0%      Mismatches: 0
Query Match:    92.7%      Indels:      0
DB:             3          Gaps:          0

US-09-606-910E-2 (1-20) x US-09-646-028-42 (1-519)
```

Qy 2 GlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
Db 241 GGTGTAACTTCTGCCCCGAGACACTCGCCAGCACCGGGTTCTACTGTCTCCGCGGCA 297

RESULT 5

US-09-646-028-46
; Sequence 46, Application US/09646028
; Patent No. 6562347
; GENERAL INFORMATION:
; APPLICANT: Kwak, Larry
; APPLICANT: Biragyn, Arva
; TITLE OF INVENTION: CHEMOKINE-TUMOR ANTIGEN FUSION PROTEINS AS CANCER VACCINES
; FILE REFERENCE: 14014.0316/P
; CURRENT APPLICATION NUMBER: US/09/646,028
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 60/077,745
; PRIOR FILING DATE: 1998-03-12
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 46
; LENGTH: 534
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of artificial sequence:/note=synthetic construct
US-09-646-028-46

Alignment Scores:
Pred. No.: 0.000742 Length: 534
Score: 101.00 Matches: 19
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 92.7% Indels: 0
Gaps: 0
DB:

US-09-606-910E-2 (1-20) x US-09-646-028-46 (1-534)

Qy 2 GlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
Db 256 GGTGTAACTTCTGCCCCGAGACACTCGCCAGCACCGGGTTCTACTGTCTCCGCGGCA 312

RESULT 6

US-10-029-517-18
; Sequence 18, Application US/10029517
; Patent No. 6716627
; GENERAL INFORMATION:
; APPLICANT: Kenneth W. Dobie
; APPLICANT: Susan J. Myers
; TITLE OF INVENTION: ANTISENSE MODULATION OF MUCIN 1, TRANSMEMBRANE EXPRESSION
; FILE REFERENCE: RTS-0352
; CURRENT APPLICATION NUMBER: US/10/029,517
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 107
; SEQ ID NO 18
; LENGTH: 572
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (67) ... (572)
US-10-029-517-18

Alignment Scores:
Pred. No.: 0.000797 Length: 572
Score: 101.00 Matches: 19
Percent Similarity: 95.0% Conservative: 0
Best Local Similarity: 95.0% Mismatches: 1
Query Match: 92.7% Indels: 0
Gaps: 0
DB:

US-09-606-910E-2 (1-20) x US-10-029-517-18 (1-572)

Qy 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
Db 499 CACGGTGTCACTCGGCCCCGAGACACACAGGCGGGGCTCCACCGCCCCCAGGCC 558

RESULT 7

US-10-029-517-101
; Sequence 101, Application US/10029517
; Patent No. 6716627
; GENERAL INFORMATION:
; APPLICANT: Kenneth W. Dobie
; APPLICANT: Susan J. Myers
; TITLE OF INVENTION: ANTISENSE MODULATION OF MUCIN 1, TRANSMEMBRANE EXPRESSION
; FILE REFERENCE: RTS-0352
; CURRENT APPLICATION NUMBER: US/10/029,517
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 107
; SEQ ID NO 101
; LENGTH: 518
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-10-029-517-101

Alignment Scores:
Pred. No.: 0.00494 Length: 518
Score: 94.00 Matches: 17
Percent Similarity: 90.0% Conservative: 1
Best Local Similarity: 85.0% Mismatches: 2
Query Match: 86.2% Indels: 0
Gaps: 0
DB:

US-09-606-910E-2 (1-20) x US-10-029-517-101 (1-518)

Qy 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
Db 423 CACGATGTCACTCAGCCCCGAGACAAACAGCAGCCCCGGGCTCCACCGCCCCCAGGCC 482

RESULT 8

US-10-029-517-102
; Sequence 102, Application US/10029517
; Patent No. 6716627
; GENERAL INFORMATION:
; APPLICANT: Kenneth W. Dobie
; APPLICANT: Susan J. Myers
; TITLE OF INVENTION: ANTISENSE MODULATION OF MUCIN 1, TRANSMEMBRANE EXPRESSION
; FILE REFERENCE: RTS-0352
; CURRENT APPLICATION NUMBER: US/10/029,517
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 107
; SEQ ID NO 102
; LENGTH: 3343
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-10-029-517-102

Alignment Scores:
Pred. No.: 0.034 Length: 3343
Score: 94.00 Matches: 17
Percent Similarity: 90.0% Conservative: 1
Best Local Similarity: 85.0% Mismatches: 2
Query Match: 86.2% Indels: 0
Gaps: 0
DB:

US-09-606-910E-2 (1-20) x US-10-029-517-102 (1-3343)

Qy 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
Db 1629 CACGATGTCACTCAGCCCCGAGACAAACAGCAGCCCCGGGCTCCACCGCCCCCAGGCC 1688

RESULT 9

US-08-479-537A-1
; Sequence 1, Application US/08479537A

Patent No. 5861381
GENERAL INFORMATION:
APPLICANT: CHAMBON, Pierre
APPLICANT: KIENY, Marie-Paule
APPLICANT: LATHE, Richard
APPLICANT: HAREUVENI, Mara
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/479,537A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
PRIOR APPLICATION NUMBER: FR 90/13101
FILING DATE: 23-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 6192 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 58..120
FEATURE:
NAME/KEY: repeat_region
LOCATION: 439..5239
OTHER INFORMATION: /note= "The nucleotides spanning
439-5239 constitute a repeated region wherein the repeat is 8
nucleotides and encodes 20 amino acids, 17 of which are fixed
and 3 which are variable."
OTHER INFORMATION: The number of such repeats varies from 1 to 80."
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 121..6166
FEATURE:
NAME/KEY: repeat_region
LOCATION: 457
OTHER INFORMATION: /note= "Nucleotide 457 is X1 = NNN
which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
or CCG, and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: repeat_region
LOCATION: 487

OTHER INFORMATION: /note= "Nucleotide 487 is Y = NNN
which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA
or ACG, and Asn = AAT or AAC."
FEATURE:
NAME/KEY: repeat_region
LOCATION: 496
OTHER INFORMATION: /note= "Nucleotide 496 is X2 = NNN
which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
or CCG, and Ala = GCT, GCC, GCA, or GCG."
US-08-479-537A-1
Alignment Scores:
Pred. No.: 0.147 Length: 6192
Score: 91.00 Matches: 17
Percent Similarity: 85.0% Conservatives: 0
Best Local Similarity: 85.0% Mismatches: 3
Query Match: 83.5% Indels: 0
DB: 2 Gaps: 0
US-09-606-910E-2 (1-20) x US-08-479-537A-1 (1-6192)
Qy 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
Db 463 CACGGTGTCACTCGGCCCGACNNNAGCGCCGNNNCCGGGCTCCACCGCCCCNNNGCC 522
RESULT 10
US-09-083-116-1
Sequence 1, Application US/09083116
Patent No. 6203795
GENERAL INFORMATION:
APPLICANT: CHAMBON, Pierre
APPLICANT: KIENY, Marie-Paule
APPLICANT: LATHE, Richard
APPLICANT: HAREUVENI, Mara
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/083,116
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/479,537
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 6192 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 58..120
FEATURE:
NAME/KEY: repeat_region
LOCATION: 439..5239
OTHER INFORMATION: /note= "The nucleotides spanning
439-5239 constitute a repeated region wherein the repeat is 8
nucleotides and encodes 20 amino acids, 17 of which are fixed
and 3 which are variable."
OTHER INFORMATION: The number of such repeats varies from 1 to 80."
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 121..6166
FEATURE:
NAME/KEY: repeat_region
LOCATION: 457
OTHER INFORMATION: /note= "Nucleotide 457 is X1 = NNN
which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
or CCG, and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: repeat_region
LOCATION: 487

SEQUENCE CHARACTERISTICS:
LENGTH: 6192 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 58..120
FEATURE:
NAME/KEY: repeat_region
LOCATION: 439..5239
OTHER INFORMATION: /note= "The nucleotides spanning 439-5239 constitute a repeated region wherein the repeat is 6
OTHER INFORMATION: nucleotides and encodes 20 amino acids, 17 of which are fixed
OTHER INFORMATION: The number of such repeats varies from 1 to 80."
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 121..6166
FEATURE:
NAME/KEY: repeat_region
LOCATION: 457
OTHER INFORMATION: /note= "Nucleotide 457 is X1 = NNN
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: repeat_region
LOCATION: 487
OTHER INFORMATION: /note= "Nucleotide 487 is Y = NNN
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA
OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
FEATURE:
NAME/KEY: repeat_region
LOCATION: 496
OTHER INFORMATION: /note= "Nucleotide 496 is X2 = NNN
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
US-09-083-116-1
Alignment Scores:
Pred. No.: 0.147 Length: 6192
Score: 91.00 Matches: 17
Percent Similarity: 85.0% Conservative: 0
Best Local Similarity: 85.0% Mismatches: 3
Query Match: 83.5% Indels: 0
DB: 3 Gaps: 0
US-09-606-910E-2 (1-20) x US-09-083-116-1 (1-6192)
QY 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
Db 463 CACGGTGTCACTCGGCCCGCCGACNNAGCCGNNCCGGGCTCCACCGCCCCCNNGCC 522
RESULT 11
US-09-134-916A-1
Sequence 1, Application US/09134916A
Patent No. 6328956
GENERAL INFORMATION:
APPLICANT: CHAMBER, Pierre
APPLICANT: KIENY, Marie-Paule
APPLICANT: LATHE, Richard
APPLICANT: HAREUVEN, Mara
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESSES:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/134,916A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/479,537
FILING DATE: 07-JUN-1995
APPLICATION NUMBER: FR 90/13101
FILING DATE: 23-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 6192 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 58..120
FEATURE:
NAME/KEY: repeat_region
LOCATION: 439..5239
OTHER INFORMATION: /note= "The nucleotides spanning 439-5239 constitute a repeated region wherein the repeat is
OTHER INFORMATION: nucleotides and encodes 20 amino acids, 17 of which are fixed
OTHER INFORMATION: The number of such repeats varies from 1 to 80."
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 121..6166
FEATURE:
NAME/KEY: repeat_region
LOCATION: 457
OTHER INFORMATION: /note= "Nucleotide 457 is X1 = NNN
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: repeat_region
LOCATION: 487
OTHER INFORMATION: /note= "Nucleotide 487 is Y = NNN
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, AC
OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
FEATURE:
NAME/KEY: repeat_region
LOCATION: 496
OTHER INFORMATION: /note= "Nucleotide 496 is X2 = NNN
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CC
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
US-09-134-916A-1
Alignment Scores:
Pred. No.: 0.147 Length: 6192
Score: 91.00 Matches: 17
Percent Similarity: 85.0% Conservative: 0
Best Local Similarity: 85.0% Mismatches: 3
Query Match: 83.5% Indels: 0
DB: 3 Gaps: 0

```
Best Local Similarity: 85.0% Mismatches: 3
Query Match: 83.5% Indels: 0
DB: 3 Gaps: 0

US-09-606-910E-2 (1-20) x US-09-134-916A-1 (1-6192)

Qy 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
Db 463 CACGGTGTCACTCGGCCCGGACNNAGCCGNNCCGGGCTCCACCGCCCCNNNGCC 522

RESULT 12
US-08-479-537A-4
; Sequence 4, Application US/08479537A
; Patent No. 5861381
; GENERAL INFORMATION:
; APPLICANT: CHAMBERON, Pierre
; APPLICANT: KIENY, Marie-Paule
; APPLICANT: LATHE, Richard
; APPLICANT: HAREUVENI, Mara
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
; TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/479,537A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 90/13101
; FILING DATE: 23-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/FR91/00835
; FILING DATE: 23-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/039,320
; FILING DATE: 04-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/403,576
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Teskin, Robin L.
; REGISTRATION NUMBER: 35,030
; REFERENCE/DOCKET NUMBER: 017753-025
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6449 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 58..120
; FEATURE:
; NAME/KEY: repeat_region
; LOCATION: 439..5239
; OTHER INFORMATION: /note= "The nucleotides spanning
; OTHER INFORMATION: 439-5239 constitute a repeated region wherein the repeat is 6
; OTHER INFORMATION: nucleotides and encodes 20 amino acids, 17 of which are fixed
```

```
; OTHER INFORMATION: The number of such repeats varies from 1 to 80."
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 121..5661
; FEATURE:
; NAME/KEY: repeat_region
; LOCATION: 457
; OTHER INFORMATION: /note= "Nucleotide 457 is X1 = NNN
; OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
; OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
; FEATURE:
; NAME/KEY: repeat_region
; LOCATION: 487
; OTHER INFORMATION: /note= "Nucleotide 487 is Y = NNN
; OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA
; OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
; FEATURE:
; NAME/KEY: repeat_region
; LOCATION: 496
; OTHER INFORMATION: /note= "Nucleotide 496 is X2 = NNN
; OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
; OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
; US-08-479-537A-4
Alignment Scores:
Pred. No.: 0.153 Length: 6449
Score: 91.00 Matches: 17
Percent Similarity: 85.0% Conservative: 0
Best Local Similarity: 85.0% Mismatches: 3
Query Match: 83.5% Indels: 0
DB: 2 Gaps: 0

US-09-606-910E-2 (1-20) x US-08-479-537A-4 (1-6449)

Qy 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
Db 463 CACGGTGTCACTCGGCCCGGACNNAGCCGNNCCGGGCTCCACCGCCCCNNNGCC 522

RESULT 13
US-09-083-116-4
; Sequence 4, Application US/09083116
; Patent No. 6203795
; GENERAL INFORMATION:
; APPLICANT: CHAMBERON, Pierre
; APPLICANT: KIENY, Marie-Paule
; APPLICANT: LATHE, Richard
; APPLICANT: HAREUVENI, Mara
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
; TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/083,116
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/479,537
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/FR91/00835
; FILING DATE: 23-OCT-1991
; PRIOR APPLICATION DATA:
```



```

; OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
; FEATURE:
; NAME/KEY: repeat_region
; LOCATION: 496
; OTHER INFORMATION: /note= "Nucleotide 496 is X2 = NNN
; OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
; OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
US-09-134-916A-4

```

```

Alignment Scores:
Pred. No.: 0.153 Length: 6449
Score: 91.00 Matches: 17
Percent Similarity: 85.0% Conservative: 0
Best Local Similarity: 85.0% Mismatches: 3
Query Match: 83.5% Indels: 0
DB: 3 Gaps: 0

```

US-09-606-910E-2 (1-20) x US-09-134-916A-4 (1-6449)

```

Qy 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
Db 463 CACGGTGTACCTCGGCCCGGACGNNAGCCGNNCCGGGCTCCACGCCGCCNNNGCC 522

```

```

RESULT 15
US-10-029-517-16
; Sequence 16, Application US/10029517
; Patent No. 6716627
; GENERAL INFORMATION:
; APPLICANT: Kenneth W. Dobie
; APPLICANT: Susan J. Myers
; TITLE OF INVENTION: ANTISENSE MODULATION OF MUCIN 1, TRANSMEMBRANE EXPRESSION
; FILE REFERENCE: RTS-0352
; CURRENT APPLICATION NUMBER: US/10/029,517
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 107
; SEQ ID NO 16
; LENGTH: 981
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: exon:exon junction
; LOCATION: (464)...(465)
; OTHER INFORMATION: exon 3b:exon 4
US-10-029-517-16

```

```

Alignment Scores:
Pred. No.: 0.0287 Length: 981
Score: 90.00 Matches: 17
Percent Similarity: 89.5% Conservative: 0
Best Local Similarity: 89.5% Mismatches: 2
Query Match: 82.6% Indels: 0
DB: 3 Gaps: 0

```

US-09-606-910E-2 (1-20) x US-10-029-517-16 (1-981)

```

Qy 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 19
Db 42 CATGGTGTACCTCGGCCCGGACGACACGCGCGCTTGGGCTCCACGCCGCCCTCCA 98

```

Search completed: April 14, 2006, 19:10:41
Job time : 118.585 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 14, 2006, 19:10:52 ; Search time 498.049 Seconds
(without alignments)
332.071 Million cell updates/sec

Title: US-09-606-910E-2

Perfect score: 109

Sequence: 1 HGVTAPDTRPAGSTAPPA 20

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n model -DEV=xlh
-Q=/abss/ABSSWEB_spool/US09606910/runat_14042006_091630_1777/app_query.fasta_1
-DB=Published_Applications NA Main -QFMT=fastap -SUFFIX=p2n.rnpbm
-MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1
-MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct
-THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext
-HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abss08
-USER=US09606910 @CEN 1 1 1364 @runat_14042006_091630_1777 -NCPU=6 -ICPU=3
-NO MMAP -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DBLEXT=7

Database :

Published Applications NA_Main:
1: /cgn2_6/prodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/prodata/1/pubpna/US08_PUBCOMB.seq.*
3: /cgn2_6/prodata/1/pubpna/US09A_PUBCOMB.seq.*
4: /cgn2_6/prodata/1/pubpna/US09B_PUBCOMB.seq.*
5: /cgn2_6/prodata/1/pubpna/US10A_PUBCOMB.seq.*
6: /cgn2_6/prodata/1/pubpna/US10B_PUBCOMB.seq.*
7: /cgn2_6/prodata/1/pubpna/US10C_PUBCOMB.seq.*
8: /cgn2_6/prodata/1/pubpna/US10D_PUBCOMB.seq.*
9: /cgn2_6/prodata/1/pubpna/US10E_PUBCOMB.seq.*
10: /cgn2_6/prodata/1/pubpna/US11_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	109	100.0	120	8	US-10-635-211-3
2	109	100.0	162	8	US-10-635-211-3
C 3	109	100.0	1424	7	US-10-447-839A-75
C 4	109	100.0	1424	8	US-10-778-859-75
5	109	100.0	1428	7	US-10-447-839A-20
6	109	100.0	1428	8	US-10-778-859-20
7	109	100.0	1527	5	US-10-057-136-19

8	109	100.0	1572	10	US-11-055-119-1	Sequence 1, Appli
9	109	100.0	1721	3	US-09-864-864-280	Sequence 280, App
10	109	100.0	1721	3	US-09-967-768A-224	Sequence 224, App
11	109	100.0	1721	5	US-10-097-340-211	Sequence 211, App
12	109	100.0	1721	5	US-10-171-311-155	Sequence 155, App
13	109	100.0	1721	6	US-10-007-926A-58	Sequence 58, Appli
14	109	100.0	1721	6	US-10-029-517-3	Sequence 3, Appli
15	109	100.0	1721	6	US-10-172-118-775	Sequence 775, App
16	109	100.0	1721	7	US-10-342-887-775	Sequence 775, App
17	109	100.0	1721	8	US-10-775-920-88	Sequence 88, Appli
18	109	100.0	1721	9	US-10-843-641A-6369	Sequence 6369, Ap
19	109	100.0	1721	9	US-10-498-704-3	Sequence 3, Appli
20	109	100.0	1721	10	US-11-050-926-211	Sequence 211, App
21	109	100.0	1761	8	US-10-635-211-1	Sequence 1, Appli
22	109	100.0	1799	7	US-10-447-839A-19	Sequence 19, Appli
23	109	100.0	1799	8	US-10-778-859-19	Sequence 19, Appli
24	109	100.0	1804	3	US-09-964-824A-573	Sequence 573, App
25	109	100.0	1804	6	US-10-029-517-17	Sequence 17, Appli
26	109	100.0	1804	7	US-10-717-597-30	Sequence 30, Appli
27	109	100.0	1804	8	US-10-775-920-84	Sequence 84, Appli
28	109	100.0	1804	9	US-10-843-641A-5876	Sequence 5876, Ap
29	109	100.0	1804	9	US-10-756-149-1275	Sequence 1275, Ap
30	109	100.0	1804	9	US-10-498-704-17	Sequence 17, Appli
31	109	100.0	1823	6	US-10-101-510-339	Sequence 339, App
32	109	100.0	2026	5	US-10-198-846-12589	Sequence 12589, A
33	109	100.0	2238	8	US-10-775-920-87	Sequence 87, Appli
34	109	100.0	2297	7	US-10-406-317-41	Sequence 41, Appli
35	109	100.0	2678	6	US-10-252-157-103	Sequence 103, App
36	109	100.0	4139	3	US-09-964-824A-105	Sequence 105, App
37	109	100.0	4139	3	US-09-964-824A-578	Sequence 578, App
38	109	100.0	4139	3	US-09-864-864-334	Sequence 334, App
39	109	100.0	4139	3	US-09-880-107-2121	Sequence 2121, App
40	109	100.0	4139	3	US-09-968-007A-751	Sequence 751, App
41	109	100.0	4139	5	US-10-171-311-157	Sequence 157, App
42	109	100.0	4139	6	US-10-177-293-310	Sequence 310, App
43	109	100.0	4139	7	US-10-440-464-155	Sequence 155, App
44	109	100.0	4139	7	US-10-734-564-53	Sequence 53, Appli
45	109	100.0	4139	8	US-10-775-920-80	Sequence 80, Appli

ALIGNMENTS

RESULT 1

US-10-635-211-3

; Sequence 3, Application US/10635211

; Publication No. US20050031649A1

; GENERAL INFORMATION:

; APPLICANT: Beijing HYDVAX Biotechnology Co. Ltd

; TITLE OF INVENTION: A recombinant fusion protein comprising BCG heat shock protein 6

; FILE REFERENCE: FPO3012US

; CURRENT FILING DATE: 2003-08-06

; NUMBER OF SEQ ID NOS: 9

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 3

; LENGTH: 120

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)..(120)

US-10-635-211-3

Alignment Scores:	4.61e-06	Length:	120
Pred. No.:	109.00	Matches:	20
Score:	100.0%	Conservative:	0
Best Local Similarity:	100.0%	Mismatches:	0
Query Match:	100.0%	Indels:	0
DB:	8	Gaps:	0

US-09-606-910E-2 (1-20) x US-10-635-211-3 (1-120)

Qy	1	HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProAla	20
Dp	22	CACGCTGTTACTCTGTCTCGGACACCCGTCGGCTCCGGGTTACCGCTCCGCCGGCT	81

```

RESULT 2
US-10-635-211-8/c
; Sequence 8, Application US/10635211
; Publication No. US20050031649A1
; GENERAL INFORMATION:
; APPLICANT: Beijing HYDVAX Biotechnology Co. Ltd
; TITLE OF INVENTION: A recombinant fusion protein comprising BCG heat shock protein 655
; TITLE OF INVENTION: and the epitope of MUC1
; FILE REFERENCE: FP03012US
; CURRENT APPLICATION NUMBER: US/10/635,211
; CURRENT FILING DATE: 2003-08-06
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 8
; LENGTH: 162
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Primer
US-10-635-211-8

```

Alignment Scores:	
Pred. No.:	5,98e-06
Score:	109.00
Percent Similarity:	100.0%
Best Local Similarity:	100.0%
Query Match:	100.0%
DB:	8
Length:	162
Matches:	20
Conservative:	0
Mismatches:	0
Indels:	0
Gaps:	0

US-09-606-910E-2 (1-20) x US-10-635-211-8 (1-162)

QY 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
pb 117 CACGGGTACCTCTGTCCGACACCCGTCCGGTCCGGTTCAACGTCGCCGGCT 58

```

RESULT 3
US-10-447-839A-75/c
; Sequence 75, Application US/10447839A
; Publication No. US20040018181A1
; GENERAL INFORMATION:
; APPLICANT: Kufe, Donald W.
; APPLICANT: Kharbanda, Sunder
; APPLICANT: Weltman, Steven D.
; TITLE OF INVENTION: MUC1 INTERFERENCE RNA COMPOSITIONS AND METHODS DERIVED THEREFROM
; FILE REFERENCE: 1000.05.009
; CURRENT APPLICATION NUMBER: US/10/447, 839A
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: 10/293, 391
; PRIOR FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 09/951, 938
; PRIOR FILING DATE: 2001-09-11
; PRIOR APPLICATION NUMBER: 60/231, 841
; PRIOR FILING DATE: 2000-09-11
; NUMBER OF SEQ ID NOS: 109
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 75
; LENGTH: 1424
; TYPE: RNA
; ORGANISM: ARTIFICIAL
; FEATURE:
; OTHER INFORMATION: Synthesized Sequence
US-10-447-839A-75

```

Alignment Scores:	3.95e-05	Length:	1424
Pred. No.:	109.00	Matches:	20
Score:	100.0%	Conservative:	0
Percent Similarity:	100.0%	Mismatches:	0
Best Local Similarity:	100.0%		

```

Query Match: 100.0% Indels: 0
DB: 7 Gaps: 0
US-09-606-910E-2 (1-20) x US-10-447-839A-75 (1-1424)

```

QY 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
|||
1020 CACGGTGTACCTCGGCCCGGGACACGAGGTCGGGCCCGGGTCCACGCGCCCCCAGCC 961
pH

RESULT 4
 US-10-778-859-75/c
 ; Sequence 75, Application US/10778859
 ; Publication No. US20050042209A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Kufe, Donald W.
 ; APPLICANT: Ohno, Tsuneya
 ; TITLE OF INVENTION: MUC1 EXTRACELLULAR DOMAIN AND CANCER TREATMENT COMPOSITIONS AND
 ; TITLE OF INVENTION: METHODS DERIVED THEREFROM
 ; FILE REFERENCE: 1000.05.011
 ; CURRENT APPLICATION NUMBER: US/10/778,859
 ; CURRENT FILING DATE: 2004-02-13
 ; PRIOR APPLICATION NUMBER: 10/293,391
 ; PRIOR FILING DATE: 2002-11-13
 ; PRIOR APPLICATION NUMBER: 10/447,839
 ; PRIOR FILING DATE: 2003-05-29
 ; PRIOR APPLICATION NUMBER: 09/951,938
 ; PRIOR FILING DATE: 2001-09-11
 ; PRIOR APPLICATION NUMBER: 60/231,841
 ; PRIOR FILING DATE: 2000-09-11
 ; NUMBER OF SEQ ID NOS: 111
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 75
 ; LENGTH: 1424
 ; TYPE: RNA
 ; ORGANISM: ARTIFICIAL
 ; FEATURE:
 ; OTHER INFORMATION: Synthesized Sequence
 ; US-10-778-859-75

Alignment Scores:		
Pred. No.:	3.95e-05	Length: 1424
Score:	109.00	Matches: 20
Percent Similarity:	100.0%	Conservative: 0
Best Local Similarity:	100.0%	Mismatches: 0
Query Match:	100.0%	Indels: 0
DB:	8	Gaps: 0

US-09-606-910E-2 (1-20) x US-10-778-859-75 (1-1424)

Qy 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
pb CACGGTGTGTCCTCGAGCCCGAGACACCAAGSCGGCCCGGGGTCCACCGCCCCCGAGCC 961
1020

```

RESULT 5
US/10/447-839A-20
; Sequence 20, Application US/10447839A
; Publication No. US20040018181A1
; GENERAL INFORMATION:
; APPLICANT: Kufe, Donald W.
; APPLICANT: Kharbanda, Suresher
; APPLICANT: Weitman, Steven D.
; TITLE OF INVENTION: MUC1 INTERFERENCE RNA
; FILE REFERENCE: 1000.05.009
; CURRENT APPLICATION NUMBER: US/10/447,839A
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: 10/7293,391
; PRIOR FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 09/951,938
; PRIOR FILING DATE: 2001-09-11
; PRIOR APPLICATION NUMBER: 60/231,841
; PRIOR FILING DATE: 2000-09-11
; NUMBER OF SEQ IDS NOS: 109
; SOFTWARE: PatentIn version 3.1.2

```

```
; SEQ ID NO 20
; LENGTH: 1428
; TYPE: RNA
; ORGANISM: RNA
US-10-447-839A-20

Alignment Scores:
Pred. No.:          3,96e-05          1428
Score:              109.00          20
Percent Similarity: 100.0%          0
Best Local Similarity: 100.0%          0
Query Match:        100.0%          0
Indels:              0
Gaps:                0
DB:                  0

US-09-606-910E-2 (1-20) x US-10-447-839A-20 (1-1428)

Qy 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
Db 406 CACGGUGUACCCUGCGCCCGGACACACAGCGCGCGGCUCCACCGCCCCCGCCAGCC 465

RESULT 6
US-10-778-859-20
; Sequence 20, Application US/10778859
; Publication No. US20050042209A1
; GENERAL INFORMATION:
; APPLICANT: Kufe, Donald W.
; TITLE OF INVENTION: MUC1 EXTRACELLULAR DOMAIN AND CANCER TREATMENT COMPOSITIONS AND
; FILE REFERENCE: 1000.05.011
; CURRENT APPLICATION NUMBER: US/10/778,859
; PRIOR FILING DATE: 2004-02-13
; PRIOR FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 10/293,391
; PRIOR FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: 10/447,839
; PRIOR FILING DATE: 2001-09-11
; PRIOR APPLICATION NUMBER: 09/951,938
; PRIOR FILING DATE: 2000-09-11
; PRIOR APPLICATION NUMBER: 60/231,841
; NUMBER OF SEQ ID NOS: 111
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 20
; LENGTH: 1428
; TYPE: RNA
; ORGANISM: RNA
US-10-778-859-20

Alignment Scores:
Pred. No.:          3,96e-05          1428
Score:              109.00          20
Percent Similarity: 100.0%          0
Best Local Similarity: 100.0%          0
Query Match:        100.0%          0
Indels:              0
Gaps:                0
DB:                  0

US-09-606-910E-2 (1-20) x US-10-778-859-20 (1-1428)

Qy 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
Db 406 CACGGUGUACCCUGCGCCCGGACACACAGCGCGCGGCUCCACCGCCCCCGCCAGCC 465

RESULT 7
US-10-057-136-19
; Sequence 19, Application US/10057136
; Publication No. US20030021770A1
; GENERAL INFORMATION:
; APPLICANT: SCHLOM, JEFFREY
; APPLICANT: KANTOR, JUDITH
; APPLICANT: KUFE, DONALD
; APPLICANT: PANICALI, DENNIS
; APPLICANT: GRITZ, LINDA
; TITLE OF INVENTION: RECOMBINANT POX VIRUS FOR IMMUNIZATION AGAINST MUC1
```

```
; TITLE OF INVENTION: TUMOR-ASSOCIATED ANTIGEN
; FILE REFERENCE: 700953/47113C
; CURRENT APPLICATION NUMBER: US/10/057,136
; CURRENT FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: 09/366,670
; PRIOR FILING DATE: 1999-08-03
; PRIOR APPLICATION NUMBER: PCT/US98/03693
; PRIOR FILING DATE: 1998-02-24
; PRIOR APPLICATION NUMBER: 60/038,253
; PRIOR FILING DATE: 1997-02-24
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
; LENGTH: 1527
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-057-136-19

Alignment Scores:
Pred. No.:          4.2e-05          1527
Score:              109.00          20
Percent Similarity: 100.0%          0
Best Local Similarity: 100.0%          0
Query Match:        100.0%          0
Indels:              0
Gaps:                0
DB:                  0

US-09-606-910E-2 (1-20) x US-10-057-136-19 (1-1527)

Qy 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
Db 247 CACGGTGTCACTCGGCCCGGACACACAGCGCGCGGCTCCACCGCCCCCGCCAGCC 306

RESULT 8
US-11-055-119-1
; Sequence 1, Application US/11055119
; Publication No. US20050142640A1
; GENERAL INFORMATION:
; APPLICANT: Taylor-Papadimitriou, Joyce
; APPLICANT: Heukamp, Lukas Carl
; APPLICANT: Offringa, Rienk
; APPLICANT: Melief, Cornelis Johanna Maria
; APPLICANT: Acres, Bruce
; APPLICANT: Thomas, Mireille
; TITLE OF INVENTION: MUC-1 derived peptides
; FILE REFERENCE: 029395-017
; CURRENT APPLICATION NUMBER: US/11/055,119
; CURRENT FILING DATE: 2005-02-11
; PRIOR APPLICATION NUMBER: US/09/658,621
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: US 60/187,215
; PRIOR FILING DATE: 2000-03-03
; PRIOR APPLICATION NUMBER: GB 9921242.5
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: EP 99 40 2237.4
; PRIOR FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1572
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (58)..(1542)
US-11-055-119-1

Alignment Scores:
Pred. No.:          4.31e-05          1572
Score:              109.00          20
Percent Similarity: 100.0%          0
Best Local Similarity: 100.0%          0
Query Match:        100.0%          0
Indels:              0
Gaps:                0
DB:                  0
```

```
US-09-606-910E-2 (1-20) x US-11-055-119-1 (1-1572)
Qy 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
Db 523 CACGGTGTCACTCGGCCCGGACACCGAGCGCGCGGCTCCACGCGCCCCCGGCC 582

RESULT 9
US-09-864-864-280
; Sequence 280, Application US/09864864
; Patent No. US20020102679A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Secrist, Heather
; APPLICANT: Lodes, Michael J.
; APPLICANT: Algate, Paul A.
; APPLICANT: Fling, Steve P.
; APPLICANT: Mannion, Jane
; APPLICANT: Benson, Darin R.
; APPLICANT: Carter, Darick
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.523
; CURRENT APPLICATION NUMBER: US/09/864,864
; CURRENT FILING DATE: 2001-05-23
; NUMBER OF SEQ ID NOS: 341
; SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 280
; LENGTH: 1721
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-864-864-280

Alignment Scores:
Pred. No.: 4,66e-05 Length: 1721
Score: 109.00 Matches: 20
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 3 Gaps: 0

US-09-606-910E-2 (1-20) x US-09-864-864-280 (1-1721)
Qy 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
Db 583 CACGGTGTCACTCGGCCCGGACACCGAGCGCGCGGCTCCACGCGCCCCCGGCC 642

RESULT 10
US-09-967-768A-224
; Sequence 224, Application US/09967768A
; Patent No. US20020150877A1
; GENERAL INFORMATION:
; APPLICANT: Augustus, Meena
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-72
; CURRENT APPLICATION NUMBER: US/09/967,768A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: US/60/236,109
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,034
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,111
; PRIOR FILING DATE: 2000-09-28
; NUMBER OF SEQ ID NOS: 325
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 224
; LENGTH: 1721
; TYPE: DNA
; ORGANISM: Homo sapiens

US-09-967-768A-224
Alignment Scores:
Pred. No.: 4,66e-05 Length: 1721
Score: 109.00 Matches: 20
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 5 Gaps: 0

US-09-606-910E-2 (1-20) x US-10-097-340-211 (1-1721)
Qy 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
Db 583 CACGGTGTCACTCGGCCCGGACACCGAGCGCGCGGCTCCACGCGCCCCCGGCC 642

RESULT 11
US-10-097-340-211
; Sequence 211, Application US/10097340
; Publication No. US20030087250A1
; GENERAL INFORMATION:
; APPLICANT: John MONAHAN
; APPLICANT: Manjula GANNAVAPURU
; APPLICANT: Sebastian HOERSCH
; APPLICANT: Shubhangi KAMATKAR
; APPLICANT: Steve G. KOVATS
; APPLICANT: Rachel E. MEYERS
; APPLICANT: Michael MORRISEY
; APPLICANT: Peter OLANDT
; APPLICANT: Ami SEN
; APPLICANT: Peter VEIBY
; APPLICANT: Gordon B. MILLS
; APPLICANT: Robert C. BAST, Jr.
; APPLICANT: Karen LU
; APPLICANT: Rosemarie SCHMANDT
; APPLICANT: Xumei ZHAO
; APPLICANT: Karen GLATT
; TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
; TITLE OF INVENTION: Assessment, Prevention, and Therapy of Ovarian Cancer
; FILE REFERENCE: MRI-030
; CURRENT APPLICATION NUMBER: US/10/097,340
; CURRENT FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: 60/276,025
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/325,149
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/276,026
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/324,967
; PRIOR FILING DATE: 2001/09/26
; PRIOR APPLICATION NUMBER: 60/311,732
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/325,102
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/323,580
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 211
; LENGTH: 1721
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-097-340-211

Alignment Scores:
Pred. No.: 4,66e-05 Length: 1721
Score: 109.00 Matches: 20
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 5 Gaps: 0
```

```

; LENGTH: 1721
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: mucin 1, transmembrane (MUC1) gene.
US-10-007-926A-58

Alignment Scores:
Pred. No.: 4.66e-05 Length: 1721
Score: 109.00 Matches: 20
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 6 Gaps: 0

US-09-606-910E-2 (1-20) x US-10-007-926A-58 (1-1721)
QY 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
Db 583 CACGGTGTACCTCGGCCCGGACACACGAGCCGGCCCGGGCTCCACCGCCCCCAGCC 642

RESULT 14
US-10-029-517-3
; Sequence 3, Application US/10029517
; Publication No. US20030148969A1
; GENERAL INFORMATION:
; APPLICANT: Kenneth W. Dobie
; APPLICANT: Susan J. Myers
; TITLE OF INVENTION: ANTISENSE MODULATION OF MUCIN 1, TRANSMEMBRANE EXPRESSION
; FILE REFERENCE: RTS-0352
; CURRENT APPLICATION NUMBER: US/10/029,517
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 107
; SEQ ID NO 3
; LENGTH: 1721
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (58)...(1605)
US-10-029-517-3

Alignment Scores:
Pred. No.: 4.66e-05 Length: 1721
Score: 109.00 Matches: 20
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 6 Gaps: 0

US-09-606-910E-2 (1-20) x US-10-029-517-3 (1-1721)
QY 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
Db 583 CACGGTGTACCTCGGCCCGGACACACGAGCCGGCCCGGGCTCCACCGCCCCCAGCC 642

RESULT 15
US-10-172-118-775
; Sequence 775, Application US/10172118
; Publication No. US20030224374A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Chris
; APPLICANT: Van 't Veer, Laura
; APPLICANT: Van de Vijver, Marc
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-175-999
; CURRENT APPLICATION NUMBER: US/10/172,118
; CURRENT FILING DATE: 2002-06-14

```

```
; PRIOR APPLICATION NUMBER: 60/380,770
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 775
; LENGTH: 1721
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NM_002456
; DATABASE ENTRY DATE: 2001-06-18
US-10-172-118-775
```

```
Alignment Scores:
Pred. No.:      4.66e-05      Length:      1721
Score:          109.00      Matches:      20
Percent Similarity: 100.0%      Conservative: 0
Best Local Similarity: 100.0%      Mismatches: 0
Query Match:      100.0%      Indels:      0
DB:              6          Gaps:      0
```

US-09-606-910E-2 (1-20) x US-10-172-118-775 (1-1721)

```
Qy      1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
Db      583 CACGGTGTCACCTCGGCCCGCCGACACACAGGCGGGCCCGGGCTCCACCGCCCCCGAGCC 642
```

Search completed: April 14, 2006, 19:35:19
Job time : 499.049 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 14, 2006, 19:18:21 ; Search time 336.098 Seconds
(without alignments)
239.682 Million cell updates/sec

Title: US-09-606-910E-2

Perfect score: 109
Sequence: 1 HGVSAPDTRPAGSTAPPA 20

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 9281099 seqs, 2013915447 residues

Total number of hits satisfying chosen parameters: 18562198

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+p2n.model -DEV=xlh
-Q=/abss/ABSSWEB.spool/US09606910/runat_14042006_091631_1823/app_query.fasta.1
-DB=Published Applications NA New -OFMT=fastap -SUFFIX=p2n.rnpbn -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=200000000 -HOST=abss08
-USER=US09606910@CGN_1_1_975@runat_14042006_091631_1823 -NCPU=6 -ICPU=3
-NO MMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA.New.*

- 1: /SIDSS/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 2: /SIDSS/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 3: /SIDSS/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 4: /SIDSS/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 5: /SIDSS/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 6: /SIDSS/ptodata/2/pubpna/US09_NEW_PUB.seq.1*
- 7: /SIDSS/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 8: /SIDSS/ptodata/2/pubpna/US10_NEW_PUB.seq.1*
- 9: /SIDSS/ptodata/2/pubpna/US10_NEW_PUB.seq.2*
- 10: /SIDSS/ptodata/2/pubpna/US10_NEW_PUB.seq.3*
- 11: /SIDSS/ptodata/2/pubpna/US11_NEW_PUB.seq.*
- 12: /SIDSS/ptodata/2/pubpna/US11_NEW_PUB.seq.2*
- 13: /SIDSS/ptodata/2/pubpna/US11_NEW_PUB.seq.3*
- 14: /SIDSS/ptodata/2/pubpna/US11_NEW_PUB.seq.4*
- 15: /SIDSS/ptodata/2/pubpna/US60_NEW_PUB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	109	100.0	328	9	US-10-517-696-41 Sequence 41, Appl
2	109	100.0	1194	9	US-10-517-696-83 Sequence 83, Appl

3	109	100.0	1378	9	US-10-517-696-64 Sequence 64, Appl
4	109	100.0	1630	9	US-10-517-696-79 Sequence 79, Appl
5	109	100.0	1634	9	US-10-517-696-60 Sequence 60, Appl
6	109	100.0	1712	9	US-10-517-696-57 Sequence 57, Appl
7	109	100.0	1721	11	US-11-233-510-19 Sequence 19, Appl
8	109	100.0	1738	9	US-10-517-696-40 Sequence 40, Appl
9	109	100.0	1755	9	US-10-517-696-44 Sequence 44, Appl
10	109	100.0	1774	7	US-10-515-872-9 Sequence 9, Appl
11	109	100.0	1803	9	US-10-517-696-70 Sequence 70, Appl
12	109	100.0	1804	9	US-10-501-035-113 Sequence 113, Appl
13	109	100.0	1808	9	US-10-517-696-77 Sequence 77, Appl
14	109	100.0	1823	9	US-10-517-696-78 Sequence 78, Appl
15	109	100.0	1874	9	US-10-517-696-59 Sequence 59, Appl
16	109	100.0	1882	9	US-10-517-696-48 Sequence 48, Appl
17	109	100.0	1918	9	US-10-517-696-43 Sequence 43, Appl
18	109	100.0	1930	9	US-10-517-696-49 Sequence 49, Appl
19	109	100.0	1945	9	US-10-517-696-47 Sequence 47, Appl
20	109	100.0	1949	9	US-10-517-696-69 Sequence 69, Appl
21	109	100.0	1953	9	US-10-517-696-39 Sequence 39, Appl
22	109	100.0	2045	9	US-10-517-696-72 Sequence 72, Appl
23	109	100.0	2049	9	US-10-517-696-53 Sequence 53, Appl
24	109	100.0	2090	9	US-10-517-696-76 Sequence 76, Appl
25	109	100.0	2094	9	US-10-517-696-52 Sequence 52, Appl
26	109	100.0	2194	9	US-10-517-696-54 Sequence 54, Appl
27	109	100.0	2255	9	US-10-517-696-38 Sequence 38, Appl
28	109	100.0	2333	9	US-10-517-696-56 Sequence 56, Appl
29	109	100.0	4139	9	US-10-826-585-39 Sequence 39, Appl
30	109	100.0	4144	9	US-10-517-696-37 Sequence 37, Appl
31	109	100.0	6245	8	US-10-401-386B-61 Sequence 61, Appl
32	94	86.2	468	8	US-10-401-386B-43 Sequence 43, Appl
33	94	86.2	921	8	US-10-401-386B-30 Sequence 30, Appl
34	94	86.2	3343	8	US-10-401-386B-32 Sequence 32, Appl
35	90	82.6	1404	7	US-10-515-872-13 Sequence 13, Appl
36	83	76.1	1030	9	US-10-517-696-42 Sequence 42, Appl
37	83	76.1	1605	9	US-10-517-696-58 Sequence 58, Appl
38	82	76.1	1798	9	US-10-517-696-50 Sequence 50, Appl
39	62	56.9	918	14	US-11-165-226-123 Sequence 123, Appl
40	59	54.1	2133	14	US-11-045-802-6 Sequence 6, Appl
41	59	54.1	3885	8	US-10-750-185-31424 Sequence 31424, A
42	58	54.1	3885	8	US-10-750-623-31424 Sequence 31424, A
C 43	58	53.2	755	14	US-11-082-154A-20 Sequence 20, Appl
C 44	57	52.3	201	14	US-11-124-368A-12374 Sequence 12374, A
C 45	57	52.3	201	14	US-11-124-368A-12377 Sequence 12377, A

ALIGNMENTS

RESULT 1
US-10-517-696-41
; Sequence 41, Application US/10517696
; Publication No. US20060051759A1
; GENERAL INFORMATION:
; APPLICANT: diadexus, Inc.
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto A.
; APPLICANT: Turner, Leah R.
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and P.
; FILE REFERENCE: DEX-0432
; CURRENT APPLICATION NUMBER: US/10/517,696
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 60/389,327
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 171
; SOFTWARE: Patent version 3.1
; SEQ ID NO 41
; LENGTH: 328
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-517-696-41
Alignment Scores:

; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 171
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 60
; LENGTH: 1634
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-517-696-60

Alignment Scores:
Pred. No.: 0.00127 Length: 1634
Score: 109.00 Matches: 20
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 9 Gaps: 0

US-09-606-910E-2 (1-20) x US-10-517-696-60 (1-1634)

Qy 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
Db 617 CACGGTGTCACTCGGCCCCGGACACAGCGCGCGGGGTCCACCGCCCCCAGCC 676

RESULT 6

US-10-517-696-57
; Sequence 57, Application US/10517696
; Publication No. US20060051759A1
; GENERAL INFORMATION:
; APPLICANT: diaDexus, Inc.
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto A.
; APPLICANT: Turner, Leah R.
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and P
; FILE REFERENCE: DEX-0432
; CURRENT APPLICATION NUMBER: US/10/517,696
; PRIOR FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 60/389,327
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 171
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 57
; LENGTH: 1712
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-517-696-57

Alignment Scores:
Pred. No.: 0.00132 Length: 1712
Score: 109.00 Matches: 20
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 9 Gaps: 0

US-09-606-910E-2 (1-20) x US-10-517-696-57 (1-1712)

Qy 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
Db 617 CACGGTGTCACTCGGCCCCGGACACAGCGCGCGGGGTCCACCGCCCCCAGCC 676

RESULT 7

US-11-233-510-19
; Sequence 19, Application US/11233510
; Publication No. US20060063190A1
; GENERAL INFORMATION:
; APPLICANT: Fischer, Timothy J.
; APPLICANT: Whitehead, Clark M.
; APPLICANT: Malinowski, Douglas P.
; APPLICANT: Marcello, Raphael
; APPLICANT: Morel, Didier
; TITLE OF INVENTION: Methods and Compositions for Evaluating

; TITLE OF INVENTION: Breast Cancer Prognosis
; FILE REFERENCE: 46143/296738
; CURRENT APPLICATION NUMBER: US/11/233,510
; CURRENT FILING DATE: 2005-09-22
; PRIOR APPLICATION NUMBER: 60/612,073
; PRIOR FILING DATE: 2004-09-22
; PRIOR APPLICATION NUMBER: 60/611,965
; PRIOR FILING DATE: 2004-09-22
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 1721
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (58)...(1605)
US-11-233-510-19

Alignment Scores:
Pred. No.: 0.00133 Length: 1721
Score: 109.00 Matches: 20
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 11 Gaps: 0

US-09-606-910E-2 (1-20) x US-11-233-510-19 (1-1721)

Qy 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
Db 583 CACGGTGTCACTCGGCCCCGGACACAGCGCGCGGGGTCCACCGCCCCCAGCC 642

RESULT 8

US-10-517-696-40
; Sequence 40, Application US/10517696
; Publication No. US20060051759A1
; GENERAL INFORMATION:
; APPLICANT: diaDexus, Inc.
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto A.
; APPLICANT: Turner, Leah R.
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and P
; FILE REFERENCE: DEX-0432
; CURRENT APPLICATION NUMBER: US/10/517,696
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 60/389,327
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 171
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 40
; LENGTH: 1738
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-517-696-40

Alignment Scores:
Pred. No.: 0.00134 Length: 1738
Score: 109.00 Matches: 20
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 9 Gaps: 0

US-09-606-910E-2 (1-20) x US-10-517-696-40 (1-1738)

Qy 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
Db 479 CACGGTGTCACTCGGCCCCGGACACAGCGCGCGGGGTCCACCGCCCCCAGCC 538

RESULT 9

```
US-10-517-696-44
; Sequence 44, Application US/10517696
; Publication No. US20060051759A1
; GENERAL INFORMATION:
; APPLICANT: diadexus, Inc.
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto A.
; APPLICANT: Turner, Leah R.
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and Pr
; FILE REFERENCE: DEX-0432
; CURRENT APPLICATION NUMBER: US/10/517,696
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 60/389,327
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 171
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 44
; LENGTH: 1755
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1682)..(1682)
; OTHER INFORMATION: n=a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1733)..(1733)
; OTHER INFORMATION: n=a, c, g, or t
US-10-517-696-44

Alignment Scores:
Pred. No.: 0.00135 Length: 1755
Score: 109.00 Matches: 20
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: Gaps: 0

US-09-606-910E-2 (1-20) x US-10-517-696-44 (1-1755)

QY 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
Db 590 CACGGTGTCACTCGCGCCCGGACACACAGCGCGCGCGGCTCCACCGCCCCCGAGCC 649

RESULT 10
US-10-515-872-9
; Sequence 9, Application US/10515872
; Publication No. US20060062798A1
; GENERAL INFORMATION:
; APPLICANT: BURDEN, NEIL
; APPLICANT: HAMBLIN, PAUL
; TITLE OF INVENTION: VACCINES
; FILE REFERENCE: PG4852
; CURRENT APPLICATION NUMBER: US/10/515,872
; CURRENT FILING DATE: 2004-11-24
; PRIOR APPLICATION NUMBER: PCT/EP03/05595
; PRIOR FILING DATE: 2003-05-23
; PRIOR APPLICATION NUMBER: GB 0212036.8
; PRIOR FILING DATE: 2002-05-24
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1774
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-515-872-9

Alignment Scores:
Pred. No.: 0.00136 Length: 1774
Score: 109.00 Matches: 20
Percent Similarity: 100.0% Conservative: 0
```

```
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: Gaps: 0

US-09-606-910E-2 (1-20) x US-10-515-872-9 (1-1774)

QY 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
Db 483 CACGGTGTCACTCGCGCCCGGACACACAGCGCGCGGCTCCACCGCCCCCGAGCC 542

RESULT 11
US-10-517-696-70
; Sequence 70, Application US/10517696
; Publication No. US20060051759A1
; GENERAL INFORMATION:
; APPLICANT: diadexus, Inc.
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto A.
; APPLICANT: Turner, Leah R.
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and Pr
; FILE REFERENCE: DEX-0432
; CURRENT APPLICATION NUMBER: US/10/517,696
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 60/389,327
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 171
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 70
; LENGTH: 1803
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-517-696-70

Alignment Scores:
Pred. No.: 0.00138 Length: 1803
Score: 109.00 Matches: 20
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: Gaps: 0

US-09-606-910E-2 (1-20) x US-10-517-696-70 (1-1803)

QY 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
Db 475 CACGGTGTCACTCGCGCCCGGACACACAGCGCGCGGCTCCACCGCCCCCGAGCC 534

RESULT 12
US-10-501-035-113
; Sequence 113, Application US/10501035
; Publication No. US20060046249A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: IDENTIFICATION OF POLYNUCLEOTIDES AND POLYPEPTIDE FOR PREDICTING
; TITLE OF INVENTION: ACTIVITY OF COMPOUNDS THAT INTERACT WITH PROTEIN TYROSINE KINASE
; TITLE OF INVENTION: AND/OR PROTEIN TYROSINE KINASE PATHWAYS
; FILE REFERENCE: D0185 PCT
; CURRENT APPLICATION NUMBER: US/10/501,035
; CURRENT FILING DATE: 2004-07-09
; PRIOR APPLICATION NUMBER: US 60/350,061
; PRIOR FILING DATE: 2002-01-18
; NUMBER OF SEQ ID NOS: 795
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 113
; LENGTH: 1804
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-501-035-113

Alignment Scores:
Pred. No.: 0.00138 Length: 1804
```

Score: 109.00 Matches: 20
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 9 Gaps: 0

US-09-606-910E-2 (1-20) x US-10-501-035-113 (1-1804)

QY 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
DB 478 CACGGTGTCACTCGGCCCGGACACACAGCGCGCGGCTCCACCGCCCCCGCCAGCC 537

RESULT 13
US-10-517-696-77
; Sequence 77, Application US/10517696
; Publication No. US20060051759A1
; GENERAL INFORMATION:
; APPLICANT: diaDexus, Inc.
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto A.
; APPLICANT: Turner, Leah R.
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and Pr
; FILE REFERENCE: DEX-0432
; CURRENT APPLICATION NUMBER: US/10/517,696
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 60/389,327
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 171
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 77
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-517-696-77

Alignment Scores:
Pred. No.: 0.00138 Length: 1808
Score: 109.00 Matches: 20
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 9 Gaps: 0

US-09-606-910E-2 (1-20) x US-10-517-696-77 (1-1808)

QY 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
DB 617 CACGGTGTCACTCGGCCCGGACACACAGCGCGGCTCCACCGCCCCCGCCAGCC 676

RESULT 14
US-10-517-696-78
; Sequence 78, Application US/10517696
; Publication No. US20060051759A1
; GENERAL INFORMATION:
; APPLICANT: diaDexus, Inc.
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto A.
; APPLICANT: Turner, Leah R.
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and Pr
; FILE REFERENCE: DEX-0432
; CURRENT APPLICATION NUMBER: US/10/517,696
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 60/389,327
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 171
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 78
; LENGTH: 1823
; TYPE: DNA

; ORGANISM: Homo sapien
US-10-517-696-78

Alignment Scores:
Pred. No.: 0.00139 Length: 1823
Score: 109.00 Matches: 20
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 9 Gaps: 0

US-09-606-910E-2 (1-20) x US-10-517-696-78 (1-1823)

QY 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
DB 617 CACGGTGTCACTCGGCCCGGACACACAGCGCGGCTCCACCGCCCCCGCCAGCC 676

RESULT 15
US-10-517-696-59
; Sequence 59, Application US/10517696
; Publication No. US20060051759A1
; GENERAL INFORMATION:
; APPLICANT: diaDexus, Inc.
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto A.
; APPLICANT: Turner, Leah R.
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and P
; FILE REFERENCE: DEX-0432
; CURRENT APPLICATION NUMBER: US/10/517,696
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 60/389,327
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 171
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 59
; LENGTH: 1874
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-517-696-59

Alignment Scores:
Pred. No.: 0.00143 Length: 1874
Score: 109.00 Matches: 20
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 9 Gaps: 0

US-09-606-910E-2 (1-20) x US-10-517-696-59 (1-1874)

QY 1 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 20
DB 617 CACGGTGTCACTCGGCCCGGACACACAGCGCGGCTCCACCGCCCCCGCCAGCC 676

Search completed: April 14, 2006, 19:46:54
Job time : 337.098 secs

This Page Blank (uspto)